AN ANALYSIS
of
Certain Psychological Tests
by the
Spearman Factor Method.

By
CHRISTOPHER JORGENSEN.

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Faculty of Philosophy, Columbia University.

London :
E. A. GOLD & CO.,
166 Blackstock Road, Highbury, N.5
1932.

ACKNOWLEDGMENT.

I have received so much assistance both directly and indirectly in connection with this study and my other work at Teachers College, Columbia University that it is not possible for me to name all those to whom my thanks and gratitude are due, and to whom they are most heartily given. I do wish to record, however, with high appreciation and deep gratitude the debt I owe to the Victorian branch of the Returned Sailors' and Soldiers' Imperial League of Australia and the Education Department of Victoria whose generosity made possible my studies abroad; to the International Institute of Teachers College, particularly Professors Wilson, Del Manzo and Kandel and the Secretary, Mrs. Reger, whose assistance in one form and another did so much to make my stay in the United States of America both profitable and enjoyable; to Professor Kitson and his assistants who so thoroughly led me through the fields of Vocational Guidance; to Professors Ruger and Pintner who, together with Professor Kitson, gave me much valuable aid in the conduct of my research work; and to Dr. A. F. Payne and Miss Edith Perry of the Personnel Bureau of the College of the City of New York whose extremely valuable assistance in putting records at my disposal and in making provision for me to administer tests greatly facilitated the collection of the data essential as the basis of the study outlined on the following pages.

CONTENTS.

CHAPTER 1.

INTRODUCTION.

(*a*) The Origin of This Study.

When one considers the large number of tests that, it is claimed, enable various kinds of human abilities or aptitudes to be measured, an inclination to be sceptical is likely to arise. Certain questions present themselves to the mind. Do these tests really measure in any marked degree the things they are supposed to measure? Where substantial correlation between test score and performance on the job can be pointed to, is this correlation due to a special relationship between the test and the job, or is the correlation principally due to some underlying general ability that is required to varying extent in all jobs and in all test performances? Or, to put this question in another form, are the correlations with criteria found for ability and aptitude tests due to the fact that each one of the tests really measures a special ability or aptitude, or does the correlation arise from the fact that a given test measures intelligence which is often one of the factors making for success on the job?

An observation made by Professor Spearman in reference to the controversy over group factors can be appropriately repeated here, as it is as true for aptitude tests as for group factors; group factors often being assumed by the aptitude tests. Spearman states, "The main cause of trouble, probably, is that current mental testing has never been built up on any general theoretical foundation. In consequence no means have been available for ascertaining how much of any correlation does derive from "g" and how much remains over to be attributed to anything else." (28, 223). For the present we can take Spearman's "g" as equivalent to what is generally meant by "general intelligence."

Then, confining attention to intelligence tests, a third question can be raised as to whether or not they all measure, as they should, substantially the same thing.

The first question, one of validity, is generally found answered to some extent by the constructors or publishers of the tests, but, apparently, very little if anything has been done in the way of supplying answers to the other two questions. This is probably due largely to the fact that a suitable technique has not been readily available. However, the Spearman method of factors can be used for these purposes, as could certain other methods that have more recently been advanced by Kelley (18), Thurstone (40), and others. In this study it was the Spearman technique that was employed.

(*b*) Purpose of This Study.

The present investigation is an attempt to analyse certain psychological tests into their general and specific factors, the general factor being something that is common to all the tests, and the specific factors being peculiar to each test, and unrelated among themselves and to the general factor.

Owing, unfortunately, to the great amount of statistical work involved in obtaining inter-correlations and tetrad differences, the present study could but scratch the surface of the vast number of tests that call for investigation. However, it was expected that some light would be thrown on the factor composition of the tests used, and also that there would emerge some evidence bearing on the advisability of further application of the Spearman method of factors to the analysis of psychological tests.

CHAPTER 2.

THE METHOD OF FACTORS.

Outline of Spearman's Theory.

Before proceeding further it would be in order to give an outline of the method of analysis that is to be used. This method is fully explained by Spearman in his book, "The Abilities of Man." Briefer accounts of it are found in Holzinger's "Résumé of the Spearman Two Factor Theory" (12), and in Chapter 18, "Psychologies of 1930," (31).

Spearman holds that all abilities, all tasks a person may perform, all tests, are composed of two factors, a general factor that is common to all, and, for each one, a specific factor which, with some few exceptions, enters into the one ability or the performance of the one task only. Thus we have the one general factor, or "g," as it is called by Spearman, running through all abilities, and an innumerable number of specific factors or "s's" —a separate one for each ability. Thus, to take some examples, if a person is set the task of finding his way out of a maze, his success will be dependent on the possession of an adequate amount of the general factor, "g," and some specific factor that is limited to finding one's way out of a maze. If a person takes a law aptitude test, his score will be a function of "g," and of the particular s involved in the law aptitude test.

The "g" and "s's" can, and in practical situations must, be looked at from two angles. They can be considered in respect to the ability or task, and they can be considered in respect to the individual in relation to his success in any ability or at any task.

Looking at the matter from the point of view of the individual, Spearman would hold that the factor "g" is possessed in but small measure by some individuals, but other fortunate people have it in large amounts. There appears to be no reason for not expecting a normal distribution of this factor among men.

Similarly, any given specific ability appears to be normally distributed, but any distribution of a specific ability is independent of the distributions of all other specific abilities and

of "g." Further, if we could measure the degree to which all, or a large number of specific abilities are possesed by a given individual, we would find, Spearman believes (28, Ch. XII.), a normal distribution. In some few things he would be a genius; in a few others, he would be woefully incompetent, while in most things he would be mediocre. It must be remembered that any success or failure achieved by the individual is dependent upon "g" as well as upon the strength of specific factors.

There really is a third factor, but this is something different, and is not included in the two-factor theory, though it is as important as either "g" or "s." It is opportunity! Unless an individual has an opportunity to find and apply those abilities he possesses in large measure, he may as well not have them at all. Further, most of us have very few of these superior abilities, and the number of things waiting for us to turn our hands to is so very large that, impelled by the necessity to eke out a livelihood at something, somehow, the chances are strongly against our finding the type of work that would allow us to use our superior abilities. This is the practical problem that has to be faced by vocational guidance workers, personnel managers, and all who would try to bring about an adjustment of human abilities and occupational requirements. And the problem becomes more acute in those activities where "s" and not "g" is of major importance.

Spearman is probably right, though we need not consider Gray entirely wrong, when he observes (28, 221) that there are not likely to be among us mute inglorious Miltons, or Cromwell guiltless of our country's blood, for to be a Milton or a Cromwell one needs to be highly possessed of "g" and such a person is likely to find his own means of rising above mediocrity. However, among the unemployed, among the discontented workers, among the inefficient workers, are many whose services would be highly valuable somewhere in some industrial or commercial organisation. At least this is Spearman's view (28, 221), and it seems quite reasonable.

If there is to be an adjustment of individual ability and job requirements, it is highly desirable that we know the maximum amount of "g" required for success in any job, and also the specific abilities involved together with their importance relative to "g." Further, we need to know the amount of "g" an individual possesses, and his strong specific abilities, or aptitudes as they are often called.

Attempts have been made to obtain information concerning both the requirements of occupations and the abilities of individuals. On the one hand we have the technique of job-analysis, and, on the other, tests of many different kinds. Of the tests two large groups concern us at present. They are (1) tests of general intelligence, and (2) special ability or aptitude tests. Clearly it would be best to keep these two groups separate, free from overlap, and, as far as possible, pure measures of what they purport to measure. But the writer has the suspicion that these desiderata are not met. He suspects that most of the tests are only to a small extent measures of what they are intended to measure. He suspects that the tests of general intelligence measure other things besides general intelligence, and that the tests of special abilities or aptitudes are partly tests of general intelligence. In this study an attempt was made to investigate for certain tests, these last two by means of the Spearman factor technique.

It will be noticed that "g" and general intelligence are here identified. Spearman does not commit himself on this point, but he does maintain that, although the theorem does not in itself supply any evidence that "g" deserves the title of "intelligence," still it does disprove the existence of any other and different general intelligence. (32).

For the purpose of simplification two further types of ability factors provided for by Spearman have so far been omitted. These factors arise from overlapping specific factors. When two abilities are very much alike, *e.g.*, reading French and taking French dictation, where a knowledge of French words helps in both performances, the specific factors are found to overlap; *i.e.*, there is more than "g" common to the two abilities. Where a small number of abilities, generally only two, are found to overlap in the manner just described, the name "group factor" is applied. These group factors are found to be not numerous. The abilities have to be very similar, and, in a number of cases where abilities have appeared to be closely related, analysis has shown that they have nothing in common above "g."

Should this overlapping of specific factors run through a number of abilities, we have what Spearman calls a "broad factor," though the name "group factor" has been applied to the same thing by others. This broad factor can be conceived of as another general factor, common to a number of abilities, but by no means all as in the case of "g." These broad factors seem to be quite few in number. Spearman and his collaborators have advanced the following:—mechanical (but a very

narrow kind involving a comprehension of the action of pulleys and levers), logical, psychological and arithmetical. Schneck claims to have proved the existence of verbal and numerical factors (27). Dr. Anastasi holds that there is a broad factor of immediate memory (4). However, acceptance of these broad factors should be held in abeyance for the present, as there is some doubt concerning the methods by which they have been obtained. Brolyer (6) has objected to the method used by Cox to establish the factor of mechanical ability, "m." In respect to Anastasi's factor of immediate memory, Holzinger (15) has said, "Excellent as the study is, the writer does not believe the evidence indicates the existence of a group factor of immediate memory." It is rather widely considered that Spearman's methods are not entirely satisfactory for the proof of the existence and the measurement of broad or group factors.

The Tetrad Difference Criterion.

The theory of one general and many specific abilities may not be greatly different from ideas held by many, long even before Spearman's time. But Spearman has put the theory on a definite, scientific basis. He has supplied a mathematical proof of the divisibility into two factors (28 Appendix i.), he has provided a method of analysis that shows whether there is any overlap of specific factors or not, and he has evolved formulae that permit the calculation of the amount of "g," and "s" or any other factor in an ability, and of the approximate strength of an individual in "g," "s" or any of the few broad factors.

The first stage in Spearman's method is the obtaining of all the inter-correlations of all the abilities under consideration. These inter-correlations are then analyzed for the purpose of ascertaining whether or not more than the one factor, "g," has been in operation to cause the correlations.

The earliest investigations had to rely on the criterion of inter-columnar correlation, but, when Spearman and Holzinger supplied a means of obtaining the probable error of the tetrad difference, the truly valid criterion could be used, and has, in fact, now been used almost exclusively for some years. The valid criterion is that of the tetrad difference tending toward zero. If any four abilities, a, b, p, q, are inter-correlated, and the difference $r_{ap}\, r_{bq} - r_{aq}\, r_{bp}$, and two similar differences be found equal to zero, the four abilities are each divisible into "g," and one specific factor which overlaps with none of the others.

When there are more than four abilities, all the possible tetrad differences must be calculated. There are 3 ${}^{n}c_{4}$ of them, and they must tend to zero. If we had correlations obtained from total populations, all our tetrad differences would be equal to zero, if there were no overlapping of specific abilities. But we never have correlations based on total populations. They are always based on samples, and, therefore, are subject to fluctuations of sampling. Consequently the tetrad differences should not be zero, but should cluster around zero in the form of a normal distribution.

Where the tetrad difference is greater than four and one half times its probable error, there is evidence of overlapping of specific factors, though even here the size of the difference may be largely due to certain disturbing influences other than chance. These disturbing influences are likely to be found when the results of a number of investigators are massed together, when there is heterogeneity of subjects, when only speed or accuracy of performance and not both is taken into account, or when there is variation of the scale used for scoring (28, 155-157; 30, 559).

CHAPTER 3.

The Spearman theory having been outlined, the way is now clear for a more definite statement of the scope of the present investigation.

The Problem.

(*a*) The main problem was to analyse by Spearman's methods certain psychological tests to the end that light might be thrown on their composition, and their meaning for vocational guidance and other purposes be increased.

(*b*) A secondary objective was to make some investigation of the value and effectiveness of the Spearman methods for the analysis of tests for the factors they contain.

In connection with (*a*) it was desired

(i) To find the extent to which "g" enters into the performance of the tasks set by the tests, and, in some cases, the extent to which it enters into the performance of the various sub-tests or sections of the tests.

(ii) To investigate the existence of overlapping or group (or broad) factors and obtain a measure of the extent to which the group factors enter into the performance of the tests.

(iii) To find the size of the specific factor that, in each case, along with "g" and any group factors, determines the quality of the performance of the task set by the test.

(iv) To compare the size of the general, the specific, and the overlapping and group factors in each test for the purpose of finding which tests or sub-tests best measure "g" and the other factors.

(v) To select from the tests and their sub-tests the distinct or non-complex abilities and find to what extent "g" and the other factors are involved in these abilities in so far as they are measured by the tests.

In connection with (*b*) it was desired to have some evidence regarding the validity and the reliability of the Spearman methods.

Expected Value of Results.

From the analysis it was expected that there would result certain findings which would be of value in the following directions :—

(1) A knowledge of the extent to which the tests analysed correlated with "g" and other factors would permit the tests to be used and interpreted with greater effectiveness.
(2) The analysis should make possible an evaluation, though not a complete one, of the tests, and give suggestions regarding the construction and use of ability tests.
(3) The study should throw light on the suitability and value of Spearman's method as a means of analysing ability tests.

Procedure.

The first step was the computation of all the inter-correlations for each group. This was carried out by the Columbia Statistical Bureau. Next, all the possible tetrad differences were obtained and compared with their probable errors. The probable error most frequently used was that given by Spearman's formula, 16A, (28, Appendix XI.). This formula gives an average value for the probable errors of the tetrad differences obtained from a group of variables. (35, 368-370). Where a more exact formula seemed desirable, Kelley's formula (18, 49) was used.

Two methods were used for the detection of overlapping specific factors. One was the tetrad difference criterion, and the other was the relationship expressed by the equation $r_{ab} = r_{ag} \cdot r_{bg}$ which needs to be approximately satisfied, if there is to be no overlap, and which henceforth in this study is referred to as the "Check."

For finding the correlation of each test or sub-test with "g," use was made of Spearman's formula (28 Appendix xvii.) :—

$$r_{ag} = \left\{ \frac{r_{ab} \cdot r_{ac} + r_{ab} \cdot r_{ad} + \text{---} + r_{ax} \cdot r_{av} + \text{---}}{r_{bc} + r_{bd} + \text{---} + r_{xy} + \text{---}} \right\}^{\frac{1}{2}}$$

In using this formula all possible units, such as $\frac{r_{ab} \cdot r_{ac}}{r_{bc}}$ were written out, and then all those units eliminated where there were any overlapping factors. When the correlations with "g" had been thus obtained, they were used in the equation $r_{ab} = r_{ag} \cdot r_{bg}$.

In some cases this check revealed overlap between tests that had been considered to be independent except for "g," and in other cases, where there had been doubt, overlap was either confirmed or shown not to be present.

After this check the correlations with "g" were again computed with the correlations of the tests now known to overlap omitted. Then the check was applied again, and the correlations of the specific factors with the total measures of ability and with each other obtained. The former were obtained by means of the formula :—

$$r_{as_a} = (1-r^2_{ag})^{\frac{1}{2}} \text{——(28 Appendix XVII).}$$

and the amount of overlap between the specific factors by the formula :—

$$r_{SaSb} = r_{ab.g} = \left\{ \frac{r_{ab} - r_{ag} \cdot r_{bg}}{(1-r^2_{ag})^{\frac{1}{2}} \ (1-r^2_{bg})^{\frac{1}{2}}} \right\} \text{28 Appendix XXII.}$$

In the case of group or broad factors, suggestions for the possible presence of such were obtained from the overlapping factors. The next step was to take those tests through which there might be a group factor, and, after partialling out "g" from the correlations and analyzing the resulting tetrad differences, to proceed as for "g." The process can be repeated for further less extensive group factors, if there is evidence (from overlap) of their possible presence.

In analysing the sub-tests, in order to have available tests known to contain "g" to serve as reference values in detecting overlap, two intelligence tests were added to each group of sub-tests. However, the reference values should be free from overlap, and, though this condition seemed to be fulfilled, when the tests were analysed along with other complete tests, they proved to overlap when grouped with the sub-tests. Consequently they could not be used to definitely determine overlap, but they did have the effect, by increasing the number and kind of tests, of adding to the reliability and validity of the measures found from the analysis.

CHAPTER 4.

THE TESTS USED.

The tests analysed were :—

Army Alpha (1925 revision). Otis—for business establishments. A.C.E. Psychological Examination.	*Intelligence Tests.*
MacQuarrie Test for Mechanical Ability. Examinations in Clerical Work (Thurstone) Teaching Aptitude (George Washington Series). Technical Information (Thurstone).	*Special Ability or Knowledge.*

In choosing these tests no special selection was made. The only reasons for their inclusion in preference to others were that (1) they are tests that actually are or have been in fairly wide use; (2) it was found more convenient to use these than other tests. There appeared to be no reason why any special selection should be made, and there was nothing to suggest that, for the purpose intended, these tests were not as good as any that might have been selected. The tests were divided into two groups, as it was not possible to give all the tests to any one group of students. These groups were :—

GROUP 1.

Army Alpha.
Otis,—for business establishments.
Teaching Aptitude.
Examination in Clerical Work.
Technical Information.
MacQuarrie Test for Mechanical Ability.

GROUP 2.

A. C. E. Psychological Examination.
Otis,—for business establishments.
Examinations in Clerical Work (each section separately timed).
Technical Information.

Sub-tests Analyzed.

From the tests used two were selected to have their sub-tests analysed. These were the MacQuarrie test for Mechanical Ability and the Thurstone Examination in Clerical Work. Difficulty, however, was experienced with the latter. Since this test is generally worked right through and the time for the whole taken, it was necessary to determine time limits for the eight sub-tests of which it is composed. The time allowed for the first sub-test (checking errors in addition and subtraction) was too long and a number of subjects "broke" the test. Apart from this the construction of this sub-test was found to be such that differences in scores did not represent corresponding differences in ability. Consequently this sub-test had to be discarded. The last sub-test (matching proverbs) had to be treated in the same way. The test was found to be too short, a number of both zero and maximum scores being found. Apart from these two sub-tests the timing decided upon proved to be quite suitable.

The MacQuarrie test for Mechanical Ability is composed of seven sub-tests.

(1) Tracing. This sub-test contains a series of vertical lines five sixteenths of an inch apart. Each line has a one sixteenth of an inch break in it. These breaks are not arranged in any regular order. The subject is required to draw as quickly as possible a pencil line through these breaks without touching the lines.

(2) Tapping. A number of circles three eights of an inch in diameter are arranged in horizontal lines and close together. The subject has to tap as rapidly as possible with the point of a pencil three times in each circle, though, if he goes outside of a circle, he is not penalized.

(3) Dotting. A series of circles five thirty seconds of an inch in diameter are placed on horizontal lines, but the distance between them varies and in no regular order. The subject is required to put one dot in each circle as quickly as possible. Dots outside the circles do not score.

(4) Copying. Twenty-five dots are arranged in the form of a three quarter inch square. A straight line pattern is given immediately to the left, and the subject is required to reproduce it by joining up the appropriate dots.

(5) LOCATION. A key consisting of a square of 36 letters is provided. Other squares of reduced dimensions are given, and in each of these squares are five dots. The subject is required to place on each dot the letter that it corresponds to in position.

(6) BLOCKS. A number of piles of nine rectangular blocks are given. These piles are built up in different ways. Five blocks in each file are marked with an X, and the subject is required to state the number of other blocks touched by each marked block.

(7) PURSUIT. From the one side of a rectangle about three by two inches ten meandering lines cross and re-cross one another until they reach the other side . The subject is required to attach the ten figures given on the one side to the ends of the corresponding lines on the other side. He is advised not to trace the lines with his pencil.

There are eight sub-tests in the Thurstone Examination in Clerical Work. The six retained for analysis may be briefly described as follows :—

Test 2. Requires the underlining of words misspelled in a passage of prose.

Test 3. Requires the letters X, Z, U and C to be struck out from among other letters arranged in haphazard fashion.

Test 4. Requires figures to be substituted for groups of four letters according to a code given at the beginning.

Test 5. Requires that from a list of men and their cities the names of the men be arranged alphabetically and according to the cities they live in.

Test 6. Requires the classification of insurance policies in three columns and according to amount, kind, and date of the policy.

Test 7. is comprised of twelve questions in arithmetical computation.

Regarding the nature of the Clerical test, Thurstone himself has said (39, 348-351), "I do not believe that office work has any special abilities that have so far been demonstrated, and hence I have confined myself to the first two criteria; namely, an appropriate intelligence level and content that appeals to the applicant for an office position. This reduces itself to the same

type of problem that we find so frequently in preparing vocational tests; namely, the preparation of an intelligence test out of relevant content." Apparently, then, the Clerical test is primarily an intelligence test. Our analysis should throw light on how well it does this job.

Validity and Reliability.

The validity of intelligence tests is unknown. There is no valid criterion. At least there is no valid criterion other than "g" which was not used in the construction of any of the tests dealt with in this study. Evidence of validity with "g" as criterion is given further on in this study.

Concerning the validity of the Test for Mechanical Ability, MacQuarrie has stated that the correlations of scores on the test with teachers ratings on mechanical ability was in no case over .48, "but different results were obtained when test scores were compared with ratings on mechanical work the students turned out. In two such studies in which most of the raters did not know who did the work they were rating, correlations of .81 were found, while in another study the correlation was .32." (21, 329).

Some evidence of the validity of the Clerical test is found in one of Thurstone's own studies. "Using as a criterion employees in an invoice company rated in five classes according to the grade of office work in which they were employed, Thurstone found a correlation of .50 between this measure of efficiency and accuracy in the tests; .42 between efficiency and speed, and .61 between efficiency and speed and accuracy combined." (36, 378).

Regarding reliability no doubts need be entertained as far as the total scores on the tests are concerned, but, before going on with the analysis, information respecting the reliability of the sub-tests of the Mechanical Ability and Clerical tests would be desirable because of the short time allowed for some of these tests. Unfortunately this information is not available for the Clerical test. The coefficients of reliability given by MacQuarrie (21, 329) for his tests are Tracing .80; Tapping .85; Dotting .74; Copying .86; Location .72; Blocks .80; Pursuit .76; and the whole test .90. These reliabilities may be considered as fairly satisfactory.

At this juncture might be mentioned certain subjective valuations of some of the tests. Kelley (19), gives the medium rating of a number of judges on a large number of tests. These ratings are for general excellence for the purpose of individual measurement and classification. Included in Kelley's list are the following tests of present concern:

Army Alpha 8, A. C. E. Psychological Test 3,

Thurstone Clerical Examination 1.

These rankings are in terms of excellence for college use. It is to be noted that the Clerical test has been given the highest possible median rating, and the Army Alpha the lowest. Does our analysis confirm or challenge these ratings?

The Subjects and Administration.

The tests were given to groups of about 160 freshmen and sophomore students at the College of the City of New York. The subjects were all male students and there was high homogeneity of race and age. Some of the tests were given by the writer, and others by responsible members of the Personnel Bureau and the Department of Education. Particular care was exercised to insure that there was no mixing of tests given by different instructors or at different times. This is the principal explanation of the fact that the number of cases is not particularly large. The objective was to avoid as far as possible all fluctuation other than that due to chance which can be checked by the use of probable errors.

CHAPTER 5.

THE ANALYSIS OF THE TESTS.

The basis of Spearman's method is the set of inter-correlation between the abilities under consideration. Consequently, it is essential that these inter-correlations be as near the true inter-correlations as possible. It has already been pointed out that effort was made to eliminate all disturbing influences except chance fluctuation. This, however, is an ideal that can never be completely attained. There always will be some attenuation, and often there are other factors, that cannot, for practical reasons, be excluded. A factor that, with some subjects, might be distinctly disturbing in a study like the present is that of age. It is seldom possible to obtain a large group of subjects who are all of the same age, but it is possible to free the inter-correlations of this influence by the partial correlation method. With younger children it might be necessary to do this, but with the group used in this study there was, according to evidence available, no need for such. The mean age of the students in the various groups was around 18 years or above, and the range was relatively small. Thus it can be said that, on the whole, the subjects had reached mental and physical maturity. This, in itself, would not have been sufficient reason for not partialling out age, but the influence of this factor had already been studied on groups of students very similar to those to whom the tests in this study were given. One of these studies was made by Schneck (27) who gives inter-correlations both without and with age partialled out, and between the two sets there is very little difference. Schneck had as subjects City College students, and City College students were the subjects in this study. Anastasi also had college students for subjects, and again the partialling out of age made very little difference. Anastasi observes (4,40) "The age correlation is in most cases very slight, ranging from .0011 to .0454 and averaging .0103." Hence, in the present study the more usual method of not partialling out age was followed.

Correction can also be made for attenuation, but this study is approached from the practical side, and the chief value of any findings is expected to be of a practical nature. What is sought is, for example, not the real correlation of any one of the tests with "g," but the value to be found in an actual testing situation. Further, testing to be most effective often requires the use of regression equations, and, in these, the raw or uncorrected coefficients are used. Consequently there seemed to be nothing to gain by correcting for attenuation, and it was not done. Perhaps because Spearman's name is connected with both

correction for attenuation and the two factor theory, it is believed by some that the latter requires the former. Such is not the case. This is specifically pointed out by Spearman himself (28 Appendix vi.), though Spearman often does use corrected coefficients. If the tetrad criterion is satisfied for corrected coefficients, it must also be satisfied for uncorrected coefficients, though the reverse statement is not necessarily true (see 4, 46).

TESTS IN GROUP 1.

The inter-correlations of the tests in the first group were found to be as shown in Table 1.

TABLE 1. Inter-correlations of Tests in Group 1.

N = 172.

Tests	1	2	3	4	5	6	Mean	Sigma
1. Army Alpha		.43	.38	.30	.29	.24	155.5	17.7
2. Teaching ...			.14	.15	.18	.01	127.3	16.1
3. Otis ...				.36	.14	.18	73.4	8.6
4. Clerical ...					.07	.12	78.5	15.2
5. Technical ...						.14	46.7	8.5
6. Mechanical							72.4	12.0

When the tetrad differences were obtained, they were found to fall fairly closely into a normal distribution, the actual median being .021 while the theoretical median or probable error was given by Spearman's formula 16A (28, xi.) as .019. There was, however, indication of overlap between the Army Alpha and the Teaching test. Two of the tetrad differences into which the correlation of these two tests entered were greater than five times the probable error; six of the eleven tetrads containing this correlation are among the highest 25 per cent. The probable error of the largest tetrad difference, .1128, by the long formula given by Kelley (18, 49) was found to be .024 which made the overlap appear less certain. The check (see p. 13), later applied, suggested overlap. Consequently in calculating the correlation of the tests with the common factor, assumed to be "g," all units containing r_{12} were thrown out. In r_{12} and elsewhere the numbers given to the tests in Table 1 are used to refer to them.

The correlations of the tests with "g" and with their specific factors are shown in Table II.

TABLE II. Correlations of Tests in Group 1 with General and Specific Factors.

Tests.	Correlation with general factor.	Correlation with specific factor.
1. Army Alpha	.73	.68
2. Teaching	.28	.96
3. Otis	.57	.82
4. Clerical	.45	.89
5. Technical	.33	.94
6. Mechanical	.29	.95

When the check was applied to the correlations with "g," the differences were all found to be small with the exception of tests 1 and 2 where the difference was .226 and the correlation of specific factors .35.

It can be concluded, then, that with one exception all the correlations given in Table 1 may be due to a common factor. That this is actually the case is not certain. The correlations are not high enough to fulfil the boundary conditions set up by Holzinger as required to be fulfilled before the existence of a common factor is definitely proved (15, 7). Again it has been assumed that this common factor is "g." It would be possible for another common factor to be present along with "g," but there is nothing apparent in the nature of the tests that suggests that this is at all likely. Further, it is just possible that tests other than the Army Alpha and the Teaching Aptitude have some overlap that our analysis has not been sensitive enough to detect or disentangle from the effects of chance.

The overlap between the Army Alpha and the Teaching Aptitude tests would appear to be not open to doubt. They overlap to the extent of .35, but, unfortunately, there is no known means of finding the probable error of this coefficient.

If it be, however, in the region of the .045 given by Yule's formula (43, 352), the correlation could scarcely have arisen from chance alone.

It is by no means easy to tell from an examination of the tests what this overlapping factor might be. The Teaching Aptitude test appears to test principally certain kinds of judgments, information, and recent memory, and these three faculties enter into the Army Alpha, but there is no warrant for concluding that this is the cause of the overlap. All that can be said is that there is something besides "g," or general intelligence, common to the two tests.

Comparison of the tests as to their effectiveness in measuring intelligence or anything else is deferred to Chapter 6.

TESTS IN GROUP 2.

The intercorrelations for the tests in Group 2 are given in Table III.

TABLE 3. Intercorrelations of Tests in Group 2.

N = 136.

Tests.	1	2	3	4	Mean	Sigma
1. A. C. E.		.59	.48	.24	195.9	38.8
2. Otis			.30	.23	67.6	11.3
3. Clerical				.03	79.6	19.5
4. Technical ...					47.9	9.4

The tetrad differences revealed no overlap of specific factors. The largest tetrad difference was found to be .0927, and its probable error given by the formula required by Kelly is .037. Thus the difference is less than three times its probable error, and so no overlap can be assumed. This does not mean, however, that no overlap is present. Nor does it necessarily follow that "g," Spearman's "g," is common to all the tests, though it probably is, and, further, there may be some common factor besides "g."

On the assumption of no overlap, the correlations with "g" (or other central factor) and the specific factors were found to be as shown in Table IV.

TABLE IV. Correlations of Tests in Group 2, with General and Specific Factors.

Tests.	Correlation with General Factor	Correlation with Specific factor.
1. A. C. E.	.98	.14
2. Otis	.71	.70
3. Clerical	.40	.92
4. Technical	.22	.97

The check was in all cases satisfied.

Comparisons of the tests and other conclusions shall be taken up in Chapter 6.

THE MACQUARRIE TEST ANALYSED.

From the point of view of analysis, the MacQuarrie Test for Mechanical Ability gave promise of being very interesting. Each section appeared to involve a definite, discernible, not very

complex ability; some of the tests appeared to have no dependence on mental power, while others seemed to be more dependent on intelligence than anything else, and still others seemed to have both intelligence and co-ordination of hand and eye; again overlapping factors seemed to be present, at least in the tests of Tapping and Dotting.

The Intercorrelations. To the seven sections of the MacQuarrie test were added two other tests known to contain "g" and between which previous analysis had shown no overlap. These two tests were the Army Alpha (1925 revision), and the Otis Test for Business Institutions. The inter-correlations obtained for all the tests are given in Table V. In Table V. and thereafter the numbers 1 to 9 are used to refer to Army Alpha, Otis, Tracing, Tapping, Dotting, Copying, Location, Blocks, and Pursuit tests respectively.

Table V. Intercorrelations of the MacQuarrie Sub-tests and Two Intelligence Tests.

		1	2	3	4	5	6	7	8	9	Mean	Sigma
1.	Army Alpha		.37	.06	.08	−.04	.21	.33	.37	.13	155.3	17.68
2.	Otis ...			.03	−.12	−.04	.18	.24	.31	.11	73.3	8.60
3.	Tracing ...				.27	.42	.36	.20	.10	.13	39.0	9.94
4.	Tapping ...					.55	.14	.17	.09	.14	44.3	7.54
5.	Dotting ...						.14	.06	.03	.25	22.3	3.70
6.	Copying ...							.46	.35	.41	47.4	15.1
7.	Location ...								.44	.30	27.3	7.79
8.	Blocks ...									.22	14.0	5.30
9.	Pursuit ...										23.6	6.09

N = 162.

The next step was to obtain the tetrad differences. If there were no overlapping factors, we would find the 378 differences normally distributed about zero, and with a median tetrad difference (signs neglected), equal to the theoretical probable error. The actual distribution was found to be far from normal, as Table VI. shows, and the observed median tetrad difference was more than double the theoretical median or probable error as obtained from Spearman's formula 16a. (28, Appendix XI.).

TABLE VI. Actual and Theoretical Distributions of Tetrad Differences.

Deviation of Tetrad Differences from Zero.	Actual Distribution.	Theoretical Distribution.
Over 11 p.e.	3	
„ 10 p.e.	1	
„ 9 p.e.	4	
„ 8 p.e.	5	
„ 7 p.e.	4	
„ 6 p.e.	4	
„ 5 p.e.	12	
„ 4 p.e.	13	1.2
„ 3 p.e.	20	6.8
„ 2 p.e.	23	25.4
„ 1 p.e.	29	61.1
0 to 1 p.e.	58	94.5
0 to —1 p.e.	47	94.5
Over —1 p.e.	45	61.1
„ —2 p.e.	24	25.4
„ —3 p.e.	30	6.8
„ —4 p.e.	30	1.2
„ —5 p.e.	9	
„ —6 p.e.	8	
„ —7 p.e.	3	
„ —8 p.e.	4	
„ —9 p.e.	2	

Actual Median Tetrad Difference = .046.
Theoretical Median Tetrad Difference = .021.

In Table VI. a theoretical distribution, based on the assumption of normal distribution is placed alongside the actual distribution obtained. The deviation from normal is quite marked, some of the differences being more than eleven times the p.e.

It was found convenient to arrange the groups of 4 tests (*e.g.*, Army Alpha, Otis, Dotting, Blocks—1, 2, 5 & 8, and having 3 tetrad differences—1258, 1285 & 1582) into three sections :—

(1) Where none of the differences exceeded 4 times the probable error, and where, consequently, no overlap could be assumed; (2) where at least one of the 3 tetrad differences for each group of 4 tests exceeded five times the probable error and where, consequently, overlapping could scarcely be doubted; and (3) where one or more of the tetrad differences was between four and five times the probable error, and where, consequently, there would be some doubt as to whether overlapping factors were indicated or not. This doubt is increased by the uncer-

tainty attached to the validity of the probable error found from the Spearman 16A. formula. There are those who object to the use of this formula, and hold that the probable error for each tetrad difference should be obtained. It was intended to find, if necessary, the separate probable errors for the differences in this section, but it was found that this was not required, as no overlapping factors were suggested whose existence had not been already revealed in section 2.

TABLE VII. shows the group of tests that fell into these three sections, and also the overlapping, or probable overlapping factors.

TABLE VII. Army Alpha, Otis and MacQuarrie Sub-tests. Overlapping, Non-overlapping and Doubtful Overlapping Factors.

No Overlap (0 — 4 p.e.)		Overlap (greater than 5 p.e.)		Doubtful (between 4 & 5 p.e.)	
Groups of Tests.			Overlapping Tests.	Groups of Tests.	Possible Overlapping Tests.
1 2 3 7	1 2 3 8	1 2 3 4	3 & 4	1 2 5 9	5 & 9
1 2 4 6	1 2 4 7	1 2 3 5	3 & 5		
1 2 4 8	1 2 4 9	1 2 3 6	3 & 6		
1 2 5 6	1 2 5 7	1 2 3 9	3 & 9		
1 2 5 8	1 2 6 8	1 2 4 5	4 & 5		
1 2 7 8	1 2 7 9	1 2 4 7	4 & 7		
1 2 8 9		1 2 6 7	6 & 7		
		1 2 6 9	6 & 9		
1 3 4 5	1 3 4 6	1 3 5 7	3 & 5	1 3 4 8	3 & 4
1 3 4 7	1 3 4 9	1 3 5 8	3 & 5	1 3 6 7	3 & 6
1 3 5 6	1 3 5 9	1 3 6 8	3 & 6	1 3 7 9	3 & 9
1 3 6 9	1 3 7 8	1 3 8 9	3 & 9		
1 4 5 9	1 4 6 7	1 4 5 6	4 & 5		
1 4 6 8	1 4 6 9	1 4 5 7	4 & 5		
1 4 7 8	1 4 7 9	1 4 5 8	4 & 5		
1 4 8 9					
1 5 6 7	1 5 6 8	1 5 8 9	5 & 9	1 5 7 9	5 & 9
1 5 6 9	1 5 7 8				
1 6 7 8	1 6 7 9	1 6 8 9	6 & 9		
1 7 8 9					
2 3 4 5	2 3 4 9	2 3 5 7	3 & 5	2 3 4 6	3 & 4
2 3 5 9	2 3 6 7	2 3 5 8	3 & 5	2 3 4 7	3 & 4
2 3 6 9	2 3 7 8	2 3 6 8	3 & 6	2 3 4 8	3 & 4
2 3 7 9				2 3 5 6	3 & 5
				2 3 8 9	3 & 9

2 4 6 9	2 4 7 9	2 4 5 6	4 & 5	2 4 5 9	4 & 5
2 4 8 9		2 4 5 7	4 & 5	2 4 6 7	none
		2 4 5 8	4 & 5	2 4 6 8	none
				very little over 4 p.e.	
		2 4 7 8	none		
			(—12 :44 — .31 :17 = .1055)		
2 5 6 7	2 5 6 8			2 5 8 9	5 & 9
2 5 6 9	2 5 7 8				
2 5 7 9					
2 6 7 8	2 6 7 9			2 6 8 9	6 & 9
2 7 8 9					
			4 & 5		3 & 7
3 4 5 8	3 4 6 8	3 4 5 6	and/or	3 4 5 7	and/or
			3 & 6		4 & 5
			4 & 5		3 & 4
3 4 6 9	3 4 7 9	3 4 5 9	and/or	3 4 6 7	and/or
			3 & 9		6 & 7
3 4 8 9		3 4 7 8	3 & 4		
			3 & 5		
3 5 8 9		3 5 6 7	and/or		
			6 & 7		
		3 5 6 8	3 & 5		
			3 & 5		
		3 5 6 9	and/or		
			6 & 9		
		3 5 7 8	3 & 5		
		3 5 7 9	3 & 5		
3 6 7 9	3 6 8 9			3 6 7 8	3 & 6
		3 7 8 9	3 & 9		
			4 & 5		
		4 5 6 7	and/or		
			6 & 7		
		4 5 6 8	4 & 5		
			4 & 5		
		4 5 6 9	and/or		
			6 & 9		
		4 5 7 8	4 & 5		
		4 5 7 9	4 & 5		
		4 5 8 9	4 & 5		
4 6 7 8	4 6 7 9				
4 6 8 9	4 7 8 9				
					5 & 9
5 6 7 8	5 6 8 9			5 6 7 9	and/or
					6 & 7
				5 7 8 9	5 & 9
6 7 8 9					

Overlapping Factors.

All possible combinations of the MacQuarrie sub-tests, taking 2 at a time, were made with the "reference values," Army Alpha and Otis (1 and 2), for the purpose of detecting overlapping factors. The result was, as seen from Table VII., that overlap was found between 3 & 4; 3 & 5; 3 & 6; 3 & 9; 4 & 5; 4 & 7; 6 & 7 and 6 & 9, with 5 & 9 doubtful. Analysis of all the other combinations of 4 tests corroborated these findings, except that of overlap between 4 & 7, and gave further indication of overlap between 5 & 9. Hence it could be concluded, tentatively at least, that there is overlap between the following tests or abilities as measured by the MacQuarrie sub-tests :—

> Tracing and Tapping, Tracing and Dotting, Tracing and Copying, Tracing and Pursuit, Tapping and Dotting, *Tapping and Location*, Copying and Location, Dotting and Pursuit, and Copying and Pursuit.

A scrutiny of the tests supports these findings, and it is conceivable that a careful comparison would have suggested such overlapping. There is one exception, however. There appears to be no reason why the tests, Tapping and Location should overlap, for no specific factor common to the two is apparent. This doubt is increased when it is noticed that evidence of this overlap appears only once. This matter shall be further investigated later on.

In Table VII. it will be noticed in column 1, headed "No Overlap," there appear together tests that in columns 2 and 3 are indicated as having overlapping factors. This is not necessarily a conflict of evidence. The explanation probably is that in column 1 chance fluctuations back towards zero tetrad difference has obscured the overlap. Thus we would expect the tests with the smallest overlap to appear least often outside of column 1, and those with the largest overlap to appear least often in column 1. Tests 4 & 7 appear only once outside column 1, and these two tests prove to have the smallest overlap, the correlation between their specific factors being .20 or less.

Now the probable error of this correlation is not known, but if we use the formula given by Yule (43, 352) for any correlation, partial or otherwise, we find the p.e. to be .05, and hence a coefficient of .20 could have arisen from chance alone. Contrasted with this is the fact that, out of a possible of twenty-two times, 4 & 5 appear in the same group only 4 times in

column 1, and in at least three of these cases a further explanation can be offered. Consequently, we would expect the specific correlation between 4 & 5 to be high. And it is the highest of all the specific correlations, being .55.

The further explanation just referred to is that in some cases there may be a common factor that is not "g" and along with which 4 & 5 do not overlap. This could be expected in groups 1 3 4 5, 2 3 4 5, and 8 3 4 5 which occur in column 1. Neither 1, 2 nor 8 overlaps with 3, 4 or 5, and, consequently, their correlation with the central factor in these groups should be zero. This proves to be the case. Actually the correlation of 1 & 2 with the factors running through 1 3 4 5 and 2 3 4 5, respectively, are both the square roots of small negative quantities, and the correlation of 8 with the factor common to 8 3 4 5 is .11, which is negligibly small.

Correlations with "g."

Is There a Factor Common to the Nine Tests?

Before an attempt was made to find the correlations of the separate tests with the central or general factor, the probability of the existence of such a central factor was investigated. It is often assumed that, when the tetrad criterion is satisfied throughout, there is a central factor, but Holzinger (15) maintains that this does not necessarily follow; the vanishing of the tetrads merely indicates that we may have a common factor. He states (15, 12) "We need to apply a boundary condition to show that we *must* have a common factor to explain the correlations as functions of non-correlated factors." Holzinger is at work on these boundary conditions. He has given (15, 7), the condition for four variables, and from it he draws the conclusion that correlations must generally be greater than 2/3 in order that a common factor be required. This makes the picture look rather black, for how are we going to get such correlations when testing college students? Does this mean that a factor that can be proven to be general to the abilities of elementary school children may not exist among college students? However, Holzinger's reminder, while suggesting caution, is not a cause for dismay. Even if the boundary condition is not met, a common factor may still be present. Further, "g" has been shown to be a factor in so many abilities, that it would seem to be more reasonable to assume its presence than its absence. In dealing with some other central factor, attention to Holzinger's boundary conditions would be more strongly demanded. And,

further still, if "g" is assumed to run through all the abilities, the abilities that do not contain it as a factor should have zero correlations with it. In the MacQuarrie sub-tests there is one test that inspection very strongly suggests is not dependent at all on "g" or "general intelligence." It is the Tapping test, and this test in the present investigation proves to have negligible correlation with "g."

The MacQuarrie tests, however, are further complicated by the presence of overlapping factors. Spearman (28 Appendix viii.) gives a method of handling these. If there are 4 abilities a, b, p and q with a & b overlapping, we can investigate whether or not there is a factor common to the 4 abilities; if we can find another ability c that does not overlap with a, b, p or q. Then we would have a factor common to a, c, p & q, and a factor common to b, c, p & q. There are (by assumption) no overlapping factors. Therefore, what is common to a, p, and q is also contained in a in one case and in b in the other. In other words, a and b have the same common factor as p and q even if they do have another (or others) as well.

Proceeding along these lines a factor can be shown to extend through the Army Alpha, the Otis, and the sub-tests of MacQuarrie. Suppose "g" is defined as the factor common to Army Alpha, Otis, and Blocks, or to tests 1, 2 and 8, to use the numbers assigned. These three are chosen because none of them overlaps with any of the others.

We have :—

g is a factor common to 1, 2 and 8.

$\therefore$ g " " " " 1 2 3 8 and to 1 2 4 8

$\therefore$ g " " " " 1 2 3 4 8

Also g " " " " 1258, to 1268, to 1278, to 1289

$\therefore$ g " " " " 1 2 3 4 5 6 7 8 9.

This method, however, is not sufficient proof. Kelley (18, 101) shows that we can have a common factor running through 4 variables, X_1, X_2, X_3, X_4, and a quite different factor through X_1, X_2, X_3, X_5. Evidence tending to support this is available from the present study. Assuming common factors in 1 3 4 5 and 1 3 4 8, the correlation of 1 with the factor common to 1 3 4 5 proved to be zero, while the correlation of 1 with the factor common to 1 3 4 8 was found to be .52. The difference is not too great to be due to chance, but it is great enough to suggest that the common factors are probably not the same. In fact one factor seems to be some form of motor ability (1 3 4 5)

and the other (1 3 4 8) intelligence. Hence in assuming a common factor we have as support only the universality of "g" and the hope that the tests not containing "g" will be discovered by zero correlation with the central factor.

METHOD OF OBTAINING CORRELATION WITH "g."

As the Army Alpha and the Otis tests are branded as intellectual tests, and as previous analysis had shown them to contain a common factor which conceivably might be "g," it was decided to take each group of four non-overlapping tests containing tests 1 and 2, to find the correlations with the central factor in each case, and then average these correlations, unless marked discrepancies challenged the validity of this procedure.

The values obtained are shown in TABLE VIII :—

TABLE VIII.—Correlations of Tests with Central Factors.

Correlations with "g." Tests 1 to 9.

Tests	1	2	3	4	5	6	7	8	9
1 2 3 7	55	43	15				73		
1 2 3 8	64	50	10					63	
1 2 4 6	79	04		0		53			
1 2 4 8	84	25		0				73	
1 2 4 9	82	0		0					38
1 2 5 6	44	38			0	56			
1 2 5 7	60	42			0		63		
1 2 5 8	59	49			0			68	
1 2 6 8	59	50				41		68	
1 2 7 8	62	49					56	66	
1 2 7 9	57	45					64		39
1 2 8 9	60	51						66	25
Aver. Correations.	.64	.37	.125	.0	.0	.50	.64	.67	.34

In ariving at the average correlations in the last line of Table VII, only a few of the groups of four tests have been taken into account, and there is the peculiarity that each group contains tests 1 and 2.

In order to obtain correlations with "g" based on the greatest number of tests, the procedure described on page 13 was used.

Its validity may be challenged, but to the writer it seems to be in order, if the assumption of a factor running through all tests is accepted, and, even if it is not, the resulting coefficients are not meaningless; they represent the ratio of the average of the correlations of a test with all other tests to the average of the correlation of the other tests among themselves.

Using this method, the coefficients given in Table IX. were obtained.

TABLE IX.—Correlations of Tests with "g."

Test.	Correlation with "g."
Army Alpha	.51
Otis	.33
Tracing	.15
Tapping	.05
Dotting	.0
Copying	.65
Location	.74
Blocks	.65
Pursuit	.36

When the check (see page 13) was applied the following probably significant differences were obtained :—

TABLE X.

"Tests not Satisfying the Check."

Tests.	Difference.
1 & 2	.2017
3 & 4	.2625
3 & 5	.4200
3 & 6	.2625
3 & 9	.2560
4 & 5	.5500
5 & 9	.2500

The rest of the differences with the possible exception of that for tests 6 and 9 (=.1660) were insignificant. To a large extent the check confirmed the findings of the tetrad analysis, but there were some important exceptions. Overlap was suggested for tests 1 and 2, for tests 4 and 7 (difference = .1330) overlap apeared very doubtful, and for tests 6 and 7 overlap was definitely contradicted, the difference being small and negative (= .1210).

This lack of perfect agreement with the findings from the tetrad analysis does not necessarily mean that there is any conflict. The failure of the tetrad analysis to show overlap between test 1 and 2, can be explained in two ways. It can be argued that the overlap was not great enough to give differences statistically significant. Another possible explanation is that overlap attributed wholly to other tests in combination with 1 and 2 should have been attributed in part to 1 and 2. This latter explanation probably fits the apparent overlap of 4 and 7, and 6 and 7. These tests occur among the overlapping ones (see Table VII) only in combination with tests known to overlap, or with tests 1 and 2. Suspicion of overlap between tests 1 and 2 is increased when it is noted that, if they do, the values given for their correlations with "g" are too high, and that, consequently, the difference (Table X) is really greater than .2017.

Hence it seems that the safest assumption would be that tests 1 and 2 overlap and that 4 and 7, and 6 and 7 do not. Greater difficulty is experienced in deciding whether or not tests 6 and 9 overlap. The difference, while less than that for tests 1 and 2, is greater than that for tests 4 and 7, and 6 and 9 occurs in significant tetrad differences not only with tests 1 and 2, but also (see Table VI.) with tests 1 and 8 and 2 and 8 which do not overlap. Consequently it was concluded that a small overlap exists between tests 6 and 9. This was considered to be the safest conclusion, as 6 and 9 would then not enter in combination into the subsequent calculations.

Correlation of Specific Factors.

Further evidence on overlaping is available from the correlations of specific factors. By means of Spearman's formula for overlapping abilities, the following correlations were obtained :—

Table XI.—Correlations of Specific Factors.

Tests.	Correlations of Specific Factors.
Army Alpha and Otis (1 & 2)... ...	.25
Tracing and Tapping (3 & 4)	.27
Tracing and Dotting (3 & 5)	.43
Tracing and Copying (3 & 6)	.36
Tracing and Pursuit (3 & 9)	.28
Tapping and Dotting (4 & 5)	.55
Tapping and Location (4 & 7)... ...	.20
Dotting and Pursuit (5 & 9)	.27
Copying and Pursuit (6 & 9)	.23

Recalculation of Correlations.

With overlap indicated between tests 1 and 2, and with such revealed as doubtful for 4 and 7 and distinctly improbable for 6 and 7, it became necessary to recalculate all correlations with "g" and between specific factors. The correlations with "g" were once again found by taking all possible combinations (28 of them) and eliminating all that contained overlapping factors. The correlations obtained are shown in Table XI which contains also the correlations given in Table VIII and Table IX and the correlations of the specific factors with the total scores on the tests. These correlations of specific factors with the total scores are given for the final set of correlations with "g" only, for it is the writer's opinion that these are the ones that should be accepted as being the most reliable.

Table XII.—Correlations of Sub-tests with Their General and Specific Factors.

		Correlations with "g."			
		1	2	3	4
	Test.	From Table VIII.	From Table IX.	Most probable values.	Correlations with specific factors *
1.	Army Alpha ...	.64	.51	.40	.92
2.	Otis	.37	.33	.21	.98
3.	Tracing ...	.125	.15	.17	.98
4.	Tapping ...	.0	.05	.14	.99
5.	Dotting ...	.0	.0	.0	1.0
6.	Copying ...	.50	.65	.68	.73
7.	Location ...	.64	.74	.80	.59
8.	Blocks	.67	.65	.67	.74
9.	Pursuit ...	.34	.36	.39	.92

* Corresponding to values in col. 3.

Using the correlations with "g" given in column 3 of Table XII, correlations between the overlapping factors were obtained. These are shown in Table XIII.

TABLE XIII.—Correlations of Specific Factors as Finally Determined.

Tests.	Correlations of Specific Factors.
Army Alpha and Otis (1 & 2)... ...	.32
Tracing and Tapping (3 & 4)	.25
Tracing and Dotting (3 & 5)	.43
Tracing and Copying (3 & 6)	.33
Tracing and Pursuit (3 & 9)	.27
Tapping and Dotting (4 & 5)	.56
Dotting and Copying (5 & 6)	.19
Dotting and Pursuit (5 & 9)	.27
Copying and Pursuit (6 & 9)	.22

Besides the values given in Table XIII, the correlations of specific factors in tests 4 and 7 and 6 and 7 were also calculated.

These were found to be 10 and—.19, respectively. Hence it is very doubtful that there is any overlap between tests 4 and 7, and there is certainly none between tests 6 and 7. A coefficient of —.19 suggests antagonism, but, as the writer understands Spearman, true negative correlation between abilities is impossible, and hence the coefficient —.19 must be considered as due to chance.

FACTORS OTHER THAN "g."

It is always possible that among a number of tests there may be factors other than "g" runnning through some of them.

A search for these group factors may be approached through an examination of the overlapping factors. Table XIV contains the overlapping factors in the nine tests under consideration.

TABLE XIV.—Overlapping Factors.

Test.	Overlap with Test.	Doubtful Overlap with Test.
1. ...	2	
2. ...	1	
3. ...	4, 5, 6 and 9	
4. ...	3 and 5	
5. ...	3, 4 and 9	... 6
6. ...	3	... 5 and 9
7.		
8.		
9. ...	3 and 5	... 6

From Table XIV it may be noticed that test 3 overlaps with tests 4, 5, 6 and nine. This suggests the possibility of a group factor running through test 3, 4, 5, 6 and 9. The first requirement for such a factor is that the five tests should all overlap pair by pair, but tests 4 and 6 and 4 and 9 have not been shown to overlap. Even if these tests do overlap the boundary conditions remain unsatisfied.

From an examination of overlap a group factor can be expected in tests 3, 4 and 5, and 3, 5, 6 and 9, but, if these are different factors, overlap should be revealed between tests 3 and 5 in each case, even after the effect of "g" has been eliminated, but the evidence is quite strong that there is no such overlap. This evidence is obtained by taking in groups of four tests 3, 4 and 5 along with every other test, and examining the tetrad differences, and by eliminating the effect of "g" from tests 3, 5, 6 and 9 and examining the tetrad differences obtained from the resulting correlations. In neither case was there any overlap between 3 and 5 revealed. It is to be noted in connection with this that "g" is not a common factor in tests 3, 4 and 5. 5 does not show any presence of it at all, and the small correlation of 4 with "g" is probably due to chance.

Consequently, it becomes necessary to reconsider the possibility of overlap between tests 4 and 6 and 4 and 9 in order to make possible a factor common to tests 3, 4, 5, 6 and 9. The correlations of the specific factors in tests 4 and 6 and 4 and 9 were found to be .16 and .12, respectively. These, on account of their smallness, had been previously neglected, but they may represent real overlap; in fact the real overlap may be greater than these figures by four or five times the probable errors. The inter-correlations between the specific factors in tests 3, 4, 5, 6 and 9 are given in Table XV.

TABLE XV.—Inter-correlations of Specific Factors in Tests 3, 4, 5, 6 and 9.

		3	4	5	6	9
3.	Tracing		.24	.43	.33	.27
4.	Tapping			.56	.16	.12
5.	Dotting				.19	.27
6.	Copying					.22
7.	Pursuit					

The tetrad differences derived from the correlations in Table XV. indicate no overlap except between tests 4 and 5, where the overlap is not to be doubted. The tetrad differences involving the specific correlations between tests 4 and 5 range from .0800 to .1373 while the probable error from Spearman's formula, 16Λ, is .023. The largest difference being only six times the probable error suggests that perhaps a more exact formular for the probable error should have been used, but the fact that all the differences involving the correlations of 4 and 5 are large tends to show that this is unnecessary, particularly as the check remained to be applied.

Calling the central factor c_1, and assuming overlap between tests 4 and 5, the correlation of each test with c_1 was found. These correlations are given in Table XVI.

TABLE XVI.—Correlations of Tests 3, 4, 5, 6 and 9. With Central Factor C_1.

Test.	Correlations with Central Factor C_1.
3. Tracing	.72
4. Tapping	.33
5. Dotting	.55
6. Copying	.45
9. Pursuit	.43

When the check was applied it was found to be satisfied in all cases except that of tests 4 and 5 where the difference was quite large. The correlation between test 4 and 5 with both "g" and c_1 eliminated was found to be as high as .48. This proves that there is in tests 4 and 5 quite a strong third factor which might be called c_2. It is impossible to determine which test, 4 or 5, is the better measure of this factor, as it is impossible to find their correlations with it. It can be noted, however, that, if tests 4 and 5 measure c_2 equally well, they will both correlate .69 (the square root of .48) with it. Henceforth in this study the correlation of both test 4 and test 5 with c_2 will be considered as .69, but this is only an approximation.

FACTOR COMPOSITION OF THE TESTS.

It does not seem possible to find any further group factors in the sub-tests of the MacQuarrie Test of Mechanical Ability. In fact the general factor "g" and the two group factors c_1 and c_2 have not been definitely proven to exist, but they are indicated, and it would seem more reasonable to accept than to

reject them, particularly as they appear to be supported by the nature of the tests. It is desirable to obtain a picture of the relative contributions of the general, broad and specific factors to the total measure of each test. In this connection the correlations as they stand cannot be summed, but their second powers can, for the sum of the squares of the correlations of each test with each central factor and its specific factor is equal to unity. This is apparent from the method of obtaining the correlation of the test with its specific factor.

TABLE XVII.—Factor Composition of Tests 1 to 9.

Test	r^2_{ag}	$r^2_{ac_1}$	$r^2_{ac_2}$	$r^2_{ag}+r^2_{ac_1}+r^2_{ac_2}$	$r^2_{as_a}$
1. Army Alpha	.16			.16	.84
2. Otis ...	.04			.04	.96
3. Tracing	.03	.52		.55	.45
4. Tapping	.02	.11	.48*	.61	.39
5. Dotting	.0	.30	.48*	.78	.22
6. Copying	.46	.20		.66	.34
7. Location	.65			.65	.35
8. Blocks ...	.45			.45	.55
9. Pursuit	.16	.18		.34	.66

* Approximations.

WHAT ABILITIES ARE THE GROUP FACTORS?

From Table XVII. it is evident that the factors isolated account to quite an appreciable extent for the total measurement of tests 3, 4, 5, 6 and 7. It may be important to determine definitely just what the abilities are that are measured by the group factors. While it is not the purpose of the present study to make this determination, some attention might be given to attempts to name or describe these factors. There does not seem to be much doubt about "g." It appears to be "g," *i.e.*, Spearman's "g," for any factor common to such differing tests as Army Alpha, Otis, Tracing, Copying, Location, Blocks and Pursuit is very unlikely to be any factor other than that which enters into all abilities, that which is what is usually meant by intelligence, "general intelligence." It is only right, however, to point out again that this factor was assumed common to all tests and not definitely proved to be so. The same reminder applies with still greater force to the group factor in tests 3, 4, 5, 6 and 9, but it would seem that the presence of these factors is more probable than not.

Factor Common to Tests 4 and 5?

It seems fairly certain that the overlapping factor c_2 in the Tapping and the Dotting tests is one of speed of muscular movement of the hand. Some degree of steadiness or precision may be involved, but it is more likely that this is contained in the group factor common to tests 3, 4, 5, 6 and 9. No mental ability or efficiency of any sensory organs seems to be involved in c_2. Either or both of these may enter into the performance of tests 4 and 5, but if so, they are probably, in these tests, of very little importance. Hence c_2 appears simply speed of hand movement of a certain type.

Factor Common to Tests 3, 4, 5, 6 and 9?

The factor common to the Tracing, Tapping, Dotting, Copying and Pursuit tests is not intelligence, and, though muscular movement is involved, it is not speed of muscular movement which appears to be what is found in test 4 and 5 alone. Speed may be involved in the performance of these tests; it almost certainly is, but it seems to be mental or dependent on sensory acuity. For instance very little movement of the hand is required in the Copying test, and the Pursuit test can be performed without any. Steadiness and precision of hand movement appear to be required in tests 3 and 5, but very little in tests 4 and 6, and not at all in test 9. Hence the factor c_1 may not be a muscular one at all. Keenness of eyesight and concentration appear to be factors in the performance of some of the tests having c_1 as a common factor, but their presence in others is not so sure, and concentration, at least, would be expected to be as potent in the Location and Blocks test. Hence it is not easy to perceive the nature of the common factor c_1. There does, however, seem to be in the five tests some vague sort of hand, eye and mental co-ordination—and, vaguer still, a sensitivity to position, direction, distance and relationships of geometrical configurations.

But very little of any of this appears to be involved in test 4, and it is noteworthy that 4 correlates less than does any of the other of the five tests with c_1, and, quite possibly, its real correlation is less than the .33 obtained. Whatever c_1 may be, it is best measured by the Tracing test, the correlation of this test with c_1 being .72, and hence the belief that c_1 is some sort of co-ordination of hand and eye is supported. However, the important thing is not what a factor is, but the part it plays in

mechanical (or other) proficiency, and this can not be found except by experimenting, though this long and laborious process might be short circuited by a knowledge of what the factor is.

Possible linkages of these factors with abilities required in occupations rather readily suggest themselves. They may be found separately or together (as in test 5) in such activities as piano playing, some forms of art, typewriting, and "assembly line" work. But these are only guesses. Not until it has been proved experimentally can we be sure that this or any other factor enters into the performance of any operation.

In attempting to isolate occupations or operations involving c_1 and c_2 as factors the Tracing and Tapping tests would perhaps be suitable and sufficient, as they overlap very little, and each has a high correlation with one of the central factors, the correlation between Tracing and c_1 being .72 and that between Tapping and c_2 being at least .69. In fact in the population at large these correlations would probably be higher as they were obtained on a homogeneous group of selected individuals.

A word of explanation is probably required here concerning this inference of the effect of homogeneity on the size of the correlation of a test with a central factor. It is well known that homogeneity tends to lower the correlations between tests. Suppose, then, that on three tests given to a heterogeneous group representative of the population as a whole the inter-correlations were all .8. It is quite conceivable that these inter-correlations might all be reduced to .5, when the tests were given to a highly homogeneous group as was done in this study. In the first case the correlation of each test with the central factor would equal—

$$\left\{\frac{.8\times.8}{.8}\right\}^{\frac{1}{2}} = .89$$

while the same correlations in the second case would be:

$$\left\{\frac{.5\times.5}{.5}\right\}^{\frac{1}{2}} = .71$$

This is not a criticism of Spearman's method, but it does show that the correlation of a test with a central factor found from college students, at least without any correction for attenuation, should be regarded as below the true correlation for the general run of individuals, particularly for mental tests, and probably, though to a less extent, for motor tests, too.

An Evaluation of the MacQuarrie Test.

The analysis that has been carried out permits of an evaluation of the MacQuarrie Test of Mechanical Ability, but, whatever the evaluation may be, it is of limited significance only. If the evaluation is favourable and the test in actual use proves to be a poor prognosticator of success in mechanical work, it is the evaluation that is at fault, as likewise it is, if it is unfavourable, and a high degree of correlation is found to exist between performance on the test and success in work requiring mechanical ability. Correlations between test scores and success in mechanical work have already been given (p. 18), but in the absence of further evidence, not much reliance can be placed on these figures as evidence of prognostic value.

If the validity of the methods used in this study be assumed, there follow certain very significant conclusions from the analysis into factors. It is the writer's oinion that the MacQuarrie Test of Mechanical Ability as a whole is not likely to be a very satisfactory measure of any ability, but it is also his opinion that there are possibilities of real value in some of the sub-tests. To begin with, all the evidence available tends to prove fairly definitely that there is no unitary factor of mechanical ability. Cox (9) using Spearman's methods unearthed an ability that he has called mechanical ability and denoted by "m," but to the present writer this is analogous to calling a differential a motor car. The connotations of the term "mechanical" are far, far wider than the recognition of the action of levers and pulleys to which Cox's "m" is confined. They range all the way from what was meant by Flavius, when of the commoners he demanded, "What, know you not that being mechanical you ought not walk upon a laboring day without the sign of your profession?" to the narrow ability, perhaps more correctly called engineering than mechanical, discovered by Cox.

The constructors of the Minnesota Tests of Mechanical Ability failed to find any general factor of mechanical ability, though they did find certain group factors among the forms of mechanical ability measured (24). Without making use of factor analysis, Kitson (20, 215-219) has given an excellent treatment of the concept of mechanical aptitude and of tests designed to measure it. He strongly doubts the existence of a factor of mechanical aptitude and the validity of mechanical aptitude tests.

Hence, if there is no such unitary factor or faculty of mechanical ability, only one type of mechanical ability, at most, can be measured by any test labelled "Test of Mechanical Ability."

But to get down closer to our analysis, the chief objection to be raised against the MacQuarrie Test is that it is a hotch-potch, and an interpretation of an obtained score is well nigh impossible. Suppose, for example, an individual makes a score somewhere near the average. What does this mean? Is he likely to meet with average success in mechanical work? This would assume a very close similarity between all types of manual work. But of more present importance than this is the question of the composition of the score. It has been shown that three of the MacQuarrie sub-tests, Copying, Location and Blocks are chiefly dependent on general intelligence; two involve some kind of motor ability, and five depend, in part, upon some other form of ability. Considering, for illustration, only two factors, intelligence and motor ability, c_2, an average score could be composed of a high score on the intelligence tests plus a low score on the motor tests, or it could be made up of a low intelligence score plus a high motor score, or, again it could be the sum of any of a large number of scores on the intelligence and motor tests. When the third factor is also given place, as it must, in the combinations, it is impossible to know what any given score means. And we need to know the composition of a score on such a test. It is not only conceivable, but probably undoubtedly true, that some kinds of mechancial work require high intelligence with or without high motor skill or anything else; and that others require high motor skill with little intelligence. Others may require a large measure of the third factor that was isolated, and so on.

SUGGESTED USE OF THE MACQUARRIE AND OTHER TESTS OF MECHANCIAL APTITUDE.

Apparently no single test is sufficient to measure any form of mechanical aptitude or ability. A battery of tests is needed, but the scores from each test must be handled scientifically; weighted proportionately to their relation to the work to be done, and not just added one to the other. Of course, the scores on the MacQuarrie sub-tests are weighted, by chance or intent, but if the weighing is good for any forms of mechanical work, it is good for those few forms only. and new systems of weighing are required for other forms of mechanical work. The psychologist, having carefully examined a given type of work for the factors involved in it, could then proceed in one of two ways. He could seek to build separate tests to measure the different mental and physical factors involved, or he could seek to construct a test

that contained all or a number of the factors in the same proportion as they are involved in the work itself. The former would probably be the easier from the angle of construction, but to use such tests would be more difficult and complicated than to use one test that "covered" a certain type of work, though the separate test would have the advantage of being applicable perhaps with certain additions and omissions, to a number of kinds of mechanical work.

If the psychologist proceeds by way of the separate tests, he must obtain the correlations between scores on his tests and performance on the work itself, and he must set up critical scores. These latter must not be overlooked, for, to take example ,whatever might be the correlation between eyesight and occupational proficiency, there will probably be found some individuals whose eyesight is inadequate. Thus a lower critical score is necessary, and, in some traits, *e.g.* intelligence, a higher critical score is required as well. Having obtained the correlations of the separate tests with work proficiency, the next important step is to convert these correlations into regression coefficients and build up a regression equation from which prediction can be made. The predictive power of the regression equation can, of course, be ascertained by finding the multiple correlation of the independent with the criterion traits. If the multiple correlation coefficient is not large, attempt should be made to improve the tests, that is to increase their correlation with the criterion while their inter-correlations are kept at a minimum.

It may be possible to use the same tests for predicting success in a number of different kinds of mechanical work. Separate regression equations would be required, and, consequently separate sets of correlations of the tests with the criteria.

To return more definitely to the MacQuarrie test, it can be stated that, instead of simply adding the scores on the sub-tests, the test could be used much more effectively if the sub-tests were aranged in two or three groups and the scores for each group correlated with some valid criterion of success. These correlations could then be thrown into a regression equation which would probably be a much better prognosticator of success than the simple total score for all the tests. Separate regression equations could be constructed for different types of mechanical work. The regression equation method is well set forth by Hull (17). However, in attempting to use this method, difficulty would be met in the MacQuarrie test, as, of the tests that measure what has in this study been called c_1, two also measure the motor factor, c_2, and three measure "g" to a greater or less extent.

One way out would be to take the 8 variables (7 MacQuarrie sub-tests plus criterion trait), and, obtaining partial correlations of the sixth order, build up a regression equation containing all the tests ungrouped. But this would be altogether too cumbersome. Another simple way out would be to use only two factors, intelligence and the motor ability in test 4 and 5, as these do not overlap. Then in one group there would be the two motor tests of Tapping and Dotting, and in the other group could be put the Copying, Location and Blocks tests which correlate fairly well with "g." There would then be but 3 variables requiring partial correlations of the 1st order only. This however would involve throwing away one of the factors measured by the test, and it may be that this factor is important in certain forms of mechanical work.

Another simple method that might prove to be the most suitable would be that of neglecting all tests except the three that measure the three factors most effectively. This would give the Location test as measuring intelligence, the Tracing test as measuring one form of motor ability, and the Tapping test as measuring the third factor. It will be recalled that the correlations of these tests with the group factors were found to be .80, .72 and .69 (approximately) respectively, and these correlations can be considered as rather satisfactory. The Tracing test, besides measuring intelligence slightly, contains another factor, but this overlaping is not greatly disturbing where there is no addition of test scores. It may have the effect, however, of reducing the partial correlations of each test with the criterion, and, consequently, of reducing the predictive value of the regression equation.

A method alternative to that of using a number of tests or groups of tests and throwing them into a regression equation, would be to find for each type of work a test that contained the same factors and in the same proportion as in the work itself. From the point of view of later administration such tests would be highly desirable, but they would be difficult to construct. The Spearman technique, however, may give some help. Certain tests are found to contain certain factors in certain proportion, and it is conceivable that it might be possible to find forms of work that contain the same factors in somewhat the same proportions. For instance, the MacQuarrie Tracing test correlates .17 with "g," and .72 with another ability. This means that 55 per cent. of the measure of the test is accounted for, and, if a type of work exists requiring principally these two factors in proportion approximating to 3 :52, the Tracing test would be a fair indicator of ability in that kind of work. Likewise it is possible that the Dotting test, containing no intelligence, but the other two in the

proportion of 30:48, might call forth the same combination of abilities as does some form of mechanical work.

The use of any of the procedures suggested here would increase the amount of work to be done by both constructor and interpreter of the tests, but the increased predictive value would make the extra effort worth while.

THE CLERICAL TEST ANALYSED.

To the sub-tests of the Thurstone Examination in Clerical Work (except the discarded first and last sub-tests) were added the Otis test for business establishments and the A.C.E. Psychological test. However, it was found that these could not be used as reference values to detect overlap in the sub-tests, as, though previous analysis had not disclosed it, they were found to overlap themselves. The inter-correlations of the two intelligence tests and the six remaining sub-tests of the clerical examination are given in Table XVIII.

TABLE XVIII. Inter-correlations of Clerical Sub-tests and Two Intelligence Tests.
N==143.

Test.	1	2	3	4	5	6	7	8	Mean*	Sigma
1. A. C. E. ...		.59	.47	.16	.22	.21	.29	.23	195.2	38.4
2. Otis ...			.24	.07	.07	.04	.26	.15	67.2	11.3
3. Spelling ...				.42	.17	.32	.13	.16	11.1	5.0
4. X, Z, U, C					.14	.17	.13	.03	10.2	6.3
5. Code ...						.27	.33	.11	12.8	5.9
6. Alphabetizing							.31	.14	17.2	4.4
7. Insurance								.06	9.9	4.3
8. Arithmetic									12.0	5.8

* N. B. The scores on the Clerical sub-tests are all in terms of the number of errors made.

In handling the tetrad differences the method used for the MacQuarrie sub-tests was again employed. That is, the groups of four tests were arranged according to the size of the largest tetrad difference in each. Those having no tetrad difference greater than 4 p.e. were put in one section; those having at least one tetrad difference greater than 5 p.e. were put in another section, and the third section was comprised of those groups of tests having one (or more) tetrad difference between 4 p.e. and 5 p.e. Those in the first section were considered as not displaying overlap of specific factors; those in the second section as containing overlapping factors, and those in the third section as being doubtful in respect of overlap. This arrangement of the groups of tests, four at a time, is found in Table XIX.

The probable error was found by Spearman's formula 16.A to be .021, which is only a little smaller than the mean (.024) of the observed tetrad differences whose distribution was found to be fairly close to normal; much closer than the distribution given in Table VI. for the MacQuarrie sub-tests, though a few differences greater than 8 times the probable error did occur. Hence some overlap confined to a few tests appeared to be present.

TABLE XIX. Overlapping Factors in Clerical Sub-tests and Two Intelligence Tests.

No overlap (0 —— 4 p.e.) Group of tests.			Overlap (greater than 5 p.e.) Group of tests.	Over-lapping tests.	Doubtful (between 4 & 5 p.e.) Groups of tests.	Possible over-lapping tests.
				1 & 2		1 & 2
1 2 3 5,	1 2 3 7,	1 2 3 8	1 2 3 4	&/or	1 2 4 6	&/or
				3 & 4		4 & 6
				1 & 2		
1 2 4 5,	1 2 4 7,	1 2 4 8	1 2 3 6	&/or		
				3 & 6		
				1 & 2		
1 2 5 8,	1 2 6 8,	1 2 7 8	1 2 5 6	&/or		
				5 & 6		
				1 & 2		
			1 2 5 7	&/or		
				5 & 7		
				1 & 2		
			1 2 6 7	&/or		
				6 & 7		
				1 & 3		1 & 7
1 3 4 5,	1 3 4 6,	1 3 4 8	1 3 5 7	&/or	1 3 4 7	&/or
				5 & 7		3 & 4
				1 & 3		1 & 3
1 3 5 8,	1 3 6 8,	1 3 7 8	1 3 6 7	&/or	1 3 5 6	&/or
				6 & 7		5 & 6
1 4 5 6,	1 4 5 7,	1 4 5 8				
1 4 6 7,	1 4 6 8,	1 4 7 8				
1 5 6 7,	1 5 6 8,	1 5 7 8				
1 6 7 8						

2345,	2346,	2347				
2348,	2356,	2357				
2358,	2367,	2368				
2378						
2456,	2457,	2458				
2467,	2468,	2478				
2567,	2568,	2578				
2678						
3458,	3468,	3478	3457	3 & 4 &/or 5 & 7	3456	3 & 4 &/or 5 & 6
			3467	3 & 4 &/or 6 & 7		
3567,	3568,	3578				
3678						
4567,	4568,	4578				
4678,	5678					

From Table XIX. it can be concluded that there is overlap between tests 1 and 2. Overlap also seems probable between tests 1 and 3, and 3 and 4, and there appears also to be a possibility of overlap between tests 5 and 6, 5 and 7, and 6 and 7. The only other test appearing in columns 2 and 3 in Table XIX. are 3 and 6, 4 and 6, and 1 and 7, but they occur only once and together with other tests where overlap is much more strongly indicated. Consequently, in the preliminary calculation of correlation with "g" the correlations between the following tests were omitted as possibly containing overlapping specific factors: 1 and 2, 1 and 3, 3 and 4, 5 and 6, 5 and 7, 6 and 7. The correlations with "g" thus obtained are shown in Table XX.

TABLE XX. Preliminary Correlations with "g." Clerical Sub-tests and Two Intelligence Tests.

Tests.	Correlations with "g."
1. A. C. E.	.69
2. Otis	.37
3. Spelling	.59
4. X, Z, U, C.	.26
5. Code	.41
6. Alphabetizing	.39
7. Insurance	.38
8. Arithmetic	.26

When the correlations given in Table XX were submitted to the check, the equation was approximately satisfied in all except four or perhaps five cases. The suspicion of overlap between tests 1 and 2, and 3 and 4 was definitely confirmed, the correlation of specific factors proving to be .49 and .35 respectively. Overlap between tests 5 and 7, and 6 and 7 still remained doubtful, the correlation of specific factors being .20 in the case of the first pair and .19 for the second. Consequently, the probability of overlap being high, though by no means certain (the probable error by Yule's formula being .054 for each) tests 5 and 7, and 6 and 7 were assumed to overlap in the re-calculation of the correlations with "g." The correlation of specific factors for tests 5 and 6 being only .13 appeared to be too low to warrant the assumption of overlap, and there was even less likelihood of overlap between tests 1 and 3. These last assumptions on the basis of overlap of specific factors are not inconsistent with anything in Table XIX. Reference to column 2 of this table will show that, assuming some overlap in tests 5 and 7, and 6 and 7, there is no need to assume overlap between tests 1 and 3. Also, with 1 and 2 overlapping, there is no need to assume that 5 and 6 overlap. Tests 5 and 6 however, do appear twice in the "doubtful overlap" column of Table XIX, but once their correlation is combined with that of tests 3 and 4, which do appear to overlap, and the remaining single case is not enough evidence to warrant the acceptance of overlap. Hence it was deemed necessary to re-calculate the correlations with "g" without omitting the correlations between tests 1 and 3, and 5 and 6. This second set of values differed only a little from the first, as can be seen from a comparison of the coefficients in Tables XX. and XXI.

TABLE XXI. Final Correlations with General and Specific Factors. Clerical Sub-tests and Two Intelligence Tests.

	Tests.	Correlations with "g."	Correlations with Specific Factors.
1.	A. C. E. ...	.66	.75
2.	Otis ...	.35	.94
3.	Spelling	.60	.79
4.	X, Z, U, C.	.26	.96
5.	Code ...	.41	.91
6.	Alphabetizing	.46	.89
7.	Insurance	.36	.93
8.	Arithmetic	.26	.96

Overlapping Factors.

As has already been mentioned, the decision was reached that the A.C.E. Psychological Examination and the Otis test for business establishments, and the Spelling checking and the X.Z.U.C. cross out contained something in common over and above "g" and that overlap was also probably present in the Code and Insurance Classification tests and in the Alphabetizing and Insurance Classification tests.

These overlapping tests, together with the extent of the overlap, are given in Table XXII.

Table XXII. Overlap of Specific Factors. Clerical Sub-tests and A.C.E. and Otis Intelligence Tests.

Overlapping Tests.	Correlations of Specific Factors.
A. C. E. and Otis	.51
Spelling and X.Z.U.C. ...	.34
Code and Insurance	.21
Alphabetizing and Insurance	.17

Group or Broad Factors.

Among the six sub-tests of the Clerical Examination and the two intelligence tests the only possibility of a factor, other than "g," common to more than two tests is in tests 5, 6 and 7, but overlap between 5 and 6 is very doubtful, and with the correlations given in Table XXI. the correlation of specific factors is reduced from .13 to .10 to which no significance can be attached. The safest conclusion seems, then, that in these tests there are no group or broad factors, only two, or at the most four, overlapping factors.

Evaluation of the Clerical Test.

The chief objection that was raised against the MacQuarrie Test for Mechanical Ability can be advanced against the Thurstone Examination in Clerical Work. This objection was that, owing to the large number of possible ways in which any given total score, particularly near the average, can be built up, interpretation of the score is rendered almost impossible. The position is perhaps not quite so bad with the Thurstone Clerical as with the MacQuarrie Mechanical Ability test, for in the former the factor "g" runs through all the sub-tests and more evenly than in the latter, and hence, as general intelligence is required in doing clerical work and in performing the clerical tests, some relationship between the two can generally be expected. However there are eight sub-tests (only six of which have been used

in this study), and it would be quite possible to omit one or two of the sub-tests and still make a fair score. This is rendered the easier by the weight given to the time taken, and the loss for amount done would be somewhat counterbalanced by the shorter time taken. But the test omitted, say the Arithmetic test, might contain something essential in some forms of clerical work, and, if the subject omitted this because he was unable to do it, he might be quite unfitted for a given clerical position, though this might not be revealed by the total score. And all this is not just the roaming of the imagination that it might seem to some. It is a fact that arithmetical computation rather frequently enters into clerical work, and it is also a fact that there are people, capable on the whole, who are markedly weak in computation. And so on for the other sub-tests.

If a sub-test measures something that enters into clerical work it is quite possible that it is essential that a certain level of ability in this thing be possessed by the applicant for a clerical position. Thus the road to measurement of clerical, as of mechanical or other ability, seems to be by way of job-analysis, critical scores and regression equations. We must find out the major abilities required by a given type of work, clerical or otherwise, devise tests that measure these abilities, set up critical scores beyond which scores cannot be considered as satisfactory, and, from the correlations of the tests with criteria of proficiency, establish regression coefficients that can be used in regression equations to give good prognostication. As an alternative to regression equations, one test may, in some cases, be used, if there can be constructed a single test which contains intelligence and the other major factors called for in the particular type of clerical work being dealt with, and if these factors are present in the test in approximately the same proportions as they are in the work.

Unless there is a strong group factor of clerical ability, one regression equation or one comprehensive test will be required for each type of clerical work. And there is no evidence of a group factor of clerical work. There is certainly none in the test that has just been analyzed, and, it will be remembered (see p. 17), Thurstone claims no more for his Examination in Clerical Work than that it is an intelligence test composed of relevant material and likely to appeal to the persons taking it. To what extent it is a test of intelligence will be dealt with further on and again in Chapter 6. The point here is that, the test not testing any unitary thing representing clerical ability, it is necessary to recourse to job analysis and the construction of tests to

measure the different abilities found to be required in the various type of clerical work. If the above arguments are sound, it follows that any test constructed as the Examination in Clerical Work is, and having the prediction made on a single total score, cannot be a good instrument for vocational guidance or selection.

It is true that the Examination in Clerical Work does correlate with proficiency on the job, but this is probably due to the fact that intelligence is a factor in the performance of both tasks. But, if it is to measure intelligence, it should measure only intelligence and not a number of other things as well, for, before any good use can be made of any test score, it is necessary to know how that score has been built up, and this is not known in the Examination in Clerical Work. The variance due to intelligence appears to be about 10 per cent. of the total variance, the remainder being due to things unknown, some or all of which may or may not correlate with Clerical ability.

Another point is to be observed. If an intelligence test is to be used to predict clerical ability, it is necessary to know whether the correlation extends along the whole range of intelligence or only along a part of it. An extremely important question is whether or not those who make the highest scores on the test are the ones most likely to be successful in clerical work. The evidence available does not support an affirmative answer to this question, yet scores of A are the ones that are regarded as being the most favourable indicators of success in clerical work. As the test measures intelligence to some extent, it follows that those who make very high scores must be of fairly high intelligence, but do people of high intelligence make the most satisfactory clerks? The answer is probably no. They may do good work for a short while, but they would be likely to become soon bored with a rather routine and monotonous work that does not give them enough mental challenge. The question of the relation of intelligence to vocational aptitude has been very well treated by Burtt (7, Ch. X.). From the data that he presents the conclusion can be drawn that persons engaged in clerical work are not of the highest intelligence, but rather slightly above average, and that the occupational turnover is very high where highly intelligent individuals are given clerical work that is not quite difficult. Hence, it appears highly probable that those who score A on the Examination in Clerical Work should not be advised to take up clerical work, unless better tests reveal that they are not particularly intelligent, or unless it is to be used as a stepping stone to an executive or other higher position.

The Examination in Clerical Work appears to be unsatisfactory in its present form as a measure of either clerical ability or intelligence. The sub-tests may be suitable for use in the measurement of clerical ability, but they must be handled by the regression equation method, and not have their scores simply added one to the other and used to predict success in clerical work of all kinds.

This conclusion is in direct contradiction to the ratings given by Kelley's judges (see p. 19). However, as neither Kelley nor his judges have given any reasons why the Examination in Clerical Work is rated so highly, the matter cannot be considered here.

CHAPTER 6.

INTERPRETATIONS AND CONCLUSIONS.

(*a*) Concerning the Tests.

The Measurement of "g" or Intelligence.

In comparing the tests used on the basis of the analysis carried out, it must be borne in mind that the comparison is in respect only to their power to measure "g." The furthest one can go is to assume that "g" in each case is the same thing and that it is general intelligence. Conceivably a test might not have a high correlation with "g," and yet be a good instrument for forecasting scholastic success. This would appear to be the case in a test in arithmetical computation to be used to foretell success in later computations of an arithmetical nature. Evidently, in this activity success is dependent more on the specific than on the general factor (correlation with "g" found to be .26 in this study).

The highest correlation with "g" and one that is strikingly high, is the .98 obtained in the second group of tests for the A.C.E. Psychological Examination. However, unrevealed overlap with the Otis test, if not with either of the other two, seems to have been partly responsible for this high correlation. When found from the inter-correlations between this test and the sub-tests of the Examination in Clerical Work, the correlation, though remaining relatively high, dropped to .66. The Otis and Army Alpha tests also gave much reduced correlations when the correlations with "g" were computed from the inter-correlations of these tests with the sub-tests of the Mechanical Ability and Clerical tests with overlapping factors omitted. The correlation of the Army Alpha with "g" dropped from .73 to .40, and the Otis from .57 to .22 in one case and from .71 to .35 in the other. What would have happened to the Clerical, the Technical Information, the Teaching Aptitude, and the Mechanical Ability tests, if they, too, had been grouped along with a larger number of tests or sub-tests is a matter for conjecture. The probability is, however, that most of these would have had their correlations with "g" raised, though it is not possible to say which ones. Therefore it can be suspected that the values obtained from Group 2 for these two tests are too low. The

reason for suspecting this is that tests that overlap (if the overlap is not detected and ruled out) have their correlations with "g" increased by the presence in the numerator of coefficients that are too high, while the other tests in the same group have these coefficients in the denominator, and hence their correlations with "g" are lowered.

It is apparent that, from the values found in this study, no very definite conclusions can be reached concerning the merits of the tests measuring intelligence. But on one point there seems to be very little doubt. The Otis test for business establishments (and there are no implications concerning any of the other Otis tests) seems to be a poor measure of general intelligence. In no case is its correlation with "g" as high as that of the Army Alpha or the A.C.E. Psychological, and these comparisons are available for different groups of tests and different groups of subjects. Even if the Spearman methods are not accepted, the fact still remains that the Otis test has rather low correlations with other tests (see Tables I., III., V., and XVIII.). The evidence is rather scanty, but what there is, from different groups of tests given to two different groups of subjects, suggests that the Examination in Clerical Work is at least as good a test of intelligence as is the Otis. But neither is a good test of intelligence. There is in the Clerical test one section (Checking Spelling—correlation with "g" .60) that is probably better than either, and there are three of the MacQuarrie sub-tests that appear to be better.

It is very difficult to compare the Army Alpha and the A.C.E. Psychological as tests of intelligence. They are not found in the same group, but a comparison based on their respective correlations with "g" suggests that the A.C.E. is the better. It correlates .98 and .66 with "g," while the corresponding correlations for the Army Alpha are .73 and .40. The other three tests, Teaching Aptitude, Technical Information, and Mechanical Ability all appear to correlate with "g" to the extent of about .30, which means that they cannot be considered as useful tests of intelligence.

From the point of view of measuring general intelligence or "g" the most interesting test is the MacQuarrie sub-test, Location, which has the relatively high correlation of .80. This is all the more surprising when it is remembered that the coefficient of reliability given for this test is only .72. Thus we have this test correlating higher with "g" than it does with itself. Consequently, it appears that the test, Location, measures little that is not "g," and, if means can be found of increasing

its reliability it should be a very good test of "g." Other tests that measure "g" fairly well are the Copying, and the Blocks in the MacQuarrie, and the Checking of Spelling Errors in the Clerical. These correlate respectively .68, .67, and .60 with "g," and, corrected for attenuation, these correlations would be somewhat higher.

Do the Tests Measure What They are Intended to Measure?

The extent to which the intelligence tests measure intelligence has just been considered, and the conclusion reached implies that the only test of definite value for this purpose is the A.C.E. Psychological Examination, and that even this test falls appreciably below what is required to measure intelligence really satisfactorily. The MacQuarrie Test for Mechanical Ability and the Thurstone Examination in Clerical Work have also been considered as measures of what is implied by the name on the face page and of what they are often used for. It was objected that no test can measure either mechanical or clerical ability, as no such unitary faculties exist, and that the tests under consideration are not likely to be satisfactory measures of any form of ability, because of the many ways in which the total score can be built up. The Teaching Aptitude Test is probably open to the same criticisms. There are no reliable criteria of success, and a total score is, again, difficult to interpret. However, this test has not been analyzed in detail in this study, and, consequently, very little can be said about it. It has a low correlation with intelligence, and overlaps with the Army Alpha test, but beyond this no information is available. The test of Technical Information undoubtedly tests technical information, but whether it adequately represents the whole field of knowledge about technical things, and what the inferences are that can be drawn from a given score are other matters outside the scope of this study which has but given some indication of the test's correlation with intelligence (about .30).

On Construction of Tests.

The methods of constructing and using ability tests indicated, in the opinion of the writer, by the analysis carried out in this study have already been outlined (see pp. 42-44 and 50, 51). The two methods advocated were the regression equation method and the single test with factor composition closely similar to that of the ability to be tested for.

In respect to the construction of intelligence tests the Spearman method appears to be very promising, and its possibility in

this direction has already been pointed out in America by Holzinger (12, 8-11). In brief, the procedure would be that of testing the tests, or better still the sections of the tests, for their correlations with "g," using in each group as large a number of tests as possible. The optimum number would appear to be about eight or nine. When the number falls below 8, the findings would be lacking in reliability, and, if more than 9 are used in one group, the increase in the number of inter-correlations and the rapid increase in the number of tetrad differences makes the computational side of the work very heavy. If the aim is to find good measures of "g" or general intelligence, the tests should be as different, one from the other, as possible in order to avoid disturbance due to overlap. Unless only one of the tests is to be used for measuring intelligence, the tests selected should be given weights proportional to their correlations with "g," *i.e.*, the regression equation method should be used.

Effect of Length of Tests.

One of the most surprising findings from the analysis of the tests used in this study is that the length ot a test has, apparently, very little effect on its power to measure "g," or, presumably, any other ability. The Location test, requiring only two minutes for the test itself, and thirty seconds for the practice exercise, is, if we assume the validity of Spearman's method, a better test of general intelligence, for certain groups of college students at least, than the Army Alpha and the Otis, which require about half an hour each, and probably better than the A.C.E. Psychological which requires an hour. Apparently, intelligence can exert itself as well in a few minutes as in so many hours.

The objection to be raised against short tests is that they do not have high reliability. The Location test has a reliability of only .72, but the other sub-tests of the MacQuarrie, some of them taking no longer than the Location, all have higher reliabilities. Reliability is a function of other things besides length of test; it is dependent upon the examiner and upon various factors in the total situation. The device of the practice exercise should, if handled to its best advantage, have the effect of increasing the reliability. In any case tests of intelligence requiring more than 10 minutes or a quarter of an hour do not seem to be called for. This statement applies to intelligence tests only; tests of ability, particularly scholastic ability, would generally need to be longer.

Equal Experience and Unfamiliar Material.

The analysis of abilities into factors throws doubt on the validity of two well-known methods of constructing intelligence tests. It has been assumed that, if individuals are tested on materials with which they are all equally familiar, the differences in the scores can be attributed to differences in intelligence. But this neglects the specific factors. In fact it would be equally true to say that the differences in scores are due to differences in the endowment of some parts of the specific factor (for, it must be remembered, the specific factor is complex). Soldiers in an army have about equal practice in rifle shooting, but can it be said that the differences in their ability to hit a target are proportional to their intelligence? The Arithmetic test in the Examination in Clerical Work has a correlation of only .26 with "g." It is therefore apparent that differences in intelligence would have only limited effect on the scores on this test, even if all subjects had had equal practice in such computations. Spearman gives (28, 242) a group factor of arithmetical ability. The amount of this factor possessed by each individual would influence his score as would the amount he has of the constituent parts of the specific factor involved.

Similar arguments apply to the claim that, if material entirely foreign to all individuals is used, the resulting differences in scores will correspond to differences in intelligence. They will to some extent, but they will also reflect differences in the amount of the requisite specific abilities possessed.

Certainly material that is entirely new to all individuals is very promising, for it eliminates the effect of different amounts of previous experience, but it is not enough to have it "new," it must also have for its performance high dependence on general intelligence. The same thing applies to material equally familiar to all; it must be highly saturated with "g."

(*b*) Concerning the Methods Used.

Though the evaluation of Spearman's methods was not the purpose of this study, opportunity was taken of seeing how some of them worked out in practice in the analysis of tests.

Consistency of Results.

Three tests, the Otis, the Technical Information, and the Clerical were in the two separate groups and given by different examiners to different subjects. If Spearman's methods are reliable, approximately the same correlations with "g" should result for these three tests in both groups. In Group 1. the

correlations with "g" were respectively .57, .33, and .45, while the corresponding coefficients from Group 2. were .71, .22, and .40 (the Clerical test in Group 2. was given in a way different from that in Group 1). The agreement is not particularly close, but the order is the same. The lack of closer agreement may be due to the presence of undetected overlapping factors. A better comparison can be expected from the correlation of the Otis with "g" when grouped with the sub-tests of the MacQuarrie, and when grouped with the sub-tests of the Clerical. In the former case the correlation was found to be .21 and in the latter .35. Again there is some consistency, but there is also some difference, and, whether this difference could or could not have been due to chance defies investigation, as the probable errors of correlations with "g" are unknown.

Another picture of the consistency of the Spearman methods is obtainable from Table VIII. It will be noticed that, with two exceptions, there is appreciable regularity of the values found for the nine tests from the different groups of variables, though this regularity could, perhaps, be partly accounted for by the presence of tests 1 and 2 in each group. Another method of comparison was to take at random groups of four non-overlapping tests and compare the correlations with "g" with the figures given in column 3 of Table XII. Such a group was that comprising tests 1 3 7 8, and the correlations with "g" were found to be .47; .18; .75; and .64, which are close to the .40; .17; .80; and .67 given in Table XI.

Consequently there is to be found a fair measure of consistency inspiring faith in Spearman's method of measuring correlation with "g." Further evidence on the reliability of Spearman's methods is given in Appendix ii.

Need for a Large Group of Tests.

When there are only four tests, marked fluctuations occur, as is seen in the correlations with "g" of tests 1 and 2 in the groups 1 2 4 6 and 1 2 4 9 in Table VIII., and in the correlations of the A.C.E. Psychological and the Otis with "g" in Group 2., compared with the same correlations when these tests were grouped with the Clerical sub-tests.

It is true that these marked fluctuations occur rather infrequently, but they do occur, and, in a given situation, it is not possible to know whether the values obtained are distorted or not by fluctuation due to chance or unrevealed overlap. Consequently, the number of tests in a group should be made as large as possible, though above nine the computations become very lengthy.

Agreement with Observation.

In some cases where overlapping or group factors were revealed, the possibility of their existence was plainly evident from observation of the tests concerned. The most conspicuous example of this is found in the Tapping and Dotting tests.

In other cases, such as between the Army Alpha and Teaching Aptitude, the Army Alpha and Otis, the A.C.E. and Otis, and the five tests having the group factor c_1 among the MacQuarrie sub-tests, the likelihood of the overlapping or group factors was not so apparent, but, at least, there was nothing tending to show that they could not be present.

The Chief Weakness.

The chief weakness in the Spearman methods is one that seems to have been generally overlooked until attention to it was called by Holzinger (15). It is the necessity of having the very exacting boundary conditions fulfilled before the presence of a factor common to all the tests of the group is known for certain.

However, in spite of their limitations the Spearman methods do appear to constitute a most powerful and greatly needed instrument for analysing psychological tests as well as other abilities, and, if such methods had been used from the inception of the testing movement, if the scientific approach could have kept in check commercialism and the grasp at novelty, there would probably be a few really good tests available to-day instead of vast numbers of doubtful value; and vocational counsellors, personal managers and others would be able to do their work more effectively.

(*c*) Concerning the Abilities Involved.

From the abilities involved in the tests a few may be selected for special mention.

(1) The ability to recognize relations in geometrical patterns appears to be highly correlated with "g." This is seen best in the Location test, but also in the Copying and Blocks tests. It is in line with Spearman's emphasis on the noegenetic processes of eduction of relations and correlates as being co-extensive with "g." (31, 362).

(2) While certain forms of purely motor ability, as manifested in the Dotting and Tapping tests, are, at least among college students, independent of "g," very little complication, as in the Tracing and Pursuit tests, is sufficient to bring this in as a factor.

(3) The ability quickly to recognize errors in spelling in a passage of prose appears to be moderately well correlated with "g" and to overlap with the rates and accuracy with which a certain few letters can be crossed out from among a large number of other letters. It is to be noted that in this spelling test most of the words were within the range of the spelling ability of most college students, but there were a few that were somewhat difficult, and that probably had the effect of complicating the abilities measured.

(4) Ability in arithmetical computation appears to be very slightly dependent on "g." The correlation of .26 found in this study agrees very closely with the correlation of .28 found for something similar by Holzinger (13, 91-97).

Implications for Vocational Guidance.

As a final summary from the point of view of vocational guidance the writer would like to stress the need for handling with great caution the many psychological tests now on the market. They may not measure to any significant extent what they are supposed to measure; general intelligence appears to enter as a factor into most of them; intelligence tests may have more in them that is specific than they do of the general factor of intelligence, and the influence of the specific factors is almost certain to be unknown and probably of very definite significance. Tests that are scored by the addition of scores made on different kinds of sub-tests should be avoided, unless there is definite proof of satisfactory correlation with an appropriate criterion.

The writer feels disposed to recommend the use of Spearman's methods by vocational guidance workers to gain further knowledge of the nature and value of the tests they use, but only those who are prepared to investigate the methods thoroughly should attempt to use them, for it is true in the case of these methods that "a little learning is a dangerous thing," and wrong or misleading conclusions can easily be drawn.

BIBLIOGRAPHY.

1. *Adams,* H. The Theory of Two Factors: An Alternative Explanation. Part I. General Factor, "G." J. App. Psych. XV. No. 1. Feb. 1931. 16-34.
2. ———The Theory of Two Factors. Part II. Specific Factors. J. App. Psych. XV. No. 4. Aug. 1931. 358-377.
3. ———A Non-intellectual Factor. J. App. Psych. Vol. XXIII. Mar. 1932. No. 3. 173-178.
4. *Anastasi.* A Group Factor in Immediate Memory. Arch. of Psych. No. 120, Oct. 1930.
5. *Asher, E. J.* The Predicative Value of Mental Tests That Satisfy Spearman's Tetrad Criterion. J. App. Psych. 1929, 13, 152-158.
6. *Brolyer, C. K.* Factors or Aptitudes Discovered or Invented. Conference on Individual Differences, New Haven, Conn. May 16, 1931. National Research Council, Washington, D.C.
7. *Burtt, H. E.* Employment Psychology, 1926, Houghton Miffin Co., N.Y.
8. *Carter, H. D.* A Reply to Professor Spearman. J. Ed. Psych. 1930, 21, 627-628.
9. *Cox, J. W.* Mechanical Aptitude. 1928, London, Methuen and Co.
10. *Dodd, S. C.* The Theory of Factors. II. Psych. Rv. 1928, 35, No. 4. 268-279.
11. ———The Theory of Factors. I. Psych. Rv. 1928, 35, No. 3. 211-234.
12. *Holzinger, K. J.* Statistical Résumé of The Spearman Two Factor Theory. Chicago Uni. Press. 1931.
13. ———On Tetrad Differences with Overlapping Variables. J. Ed. Psych. 1929, 20, No. 2. 91-97.
14. ———An Application of Spearman's Methods to the Construction of Intelligence Tests. Conference on Individual Differences, Washington, D.C. May 9 and 10, 1930. National Research Council, Washington, D.C.
15. ———On Factor Theory. Conference on Individual Differences, New Haven, Conn. May 16, 1931. National Research Council, Washington, D.C.

16. ———and *Swineford, F.* Uniqueness of Factor Patterns. J. Ed. Psych. Vol. XXIII., April 1932. No. 4. 247-258.

17. *Hull, C. L.* Autitude Testing, Yonkers-on-Hudson : World Book Co., 1928.

18. *Kelley, T. L.* Crossroads in the Mind of Man, 1928, Stanford Uni. Press.

19. ———Interpretation of Educational Measurements. 1927, World Book Co.

20. *Kitson, H. D.* Does General Mechanical Aptitude Exist? Ind. Arts and Voc. Ed. June 1930, 215-219.

21. *MacQuarrie, T. W.* A Mechanical Ability Test. J. Per. Res. 5, No. 9. 1926-27, 329-337.

22. *Melli, R.* Researches Sur Les Formes d'Intelligence, Arch. de Psych. 1930, 22, 201-284.

23. ———A Propos de la Théorie des Facteurs; Réponse à M. C. Spearman. Arch. de Psych. 1930, 22, 328-332.

24. *Paterson, Elliott, Anderson, Toops, Heidbreder.* Minnesota Mechanical Ability Tests. 1930, Uni. of Minnesota Press.

25. *Pearson, K. and Moul, M.* The Mathematics of Intelligence. I. The Sampling Errors in the Theory of a Generalized Factor, Biometrika, 1927, 19, 246-291.

26.—*Peatman, J. G.* Factors measured by the Thorndike Intelligence Examination. Arch. of Psych. 1931.

27. *Schneck, M. M. R.* The Measurement of Verbal and Numerical Abilities. Arch. of Psych. No. 107, 1929.

28. *Spearman, C.* The Abilities of Man, 1927. N.Y. McMillan.

29. ———The Uniqueness of "G." J. Ed. Psych. No. 20, 1929, 212-

30. ———Disturbers of Tetrad Differences, J. Ed. Psych. No. 21, 1930, 559-

31. ———"G" and After,—A School to End Schools. Psychologies of 1930, Clark Uni. Press, 1930, 339-366.

32. ———What the Theories of Factors is not. J. Ed. Psych. Vol. 22, No. 2. 112-117.

33. ———A Truce to Barking In. J. Ed. Psych. 1930, 21, 110-

34. ———La Théorie des Facteurs. Arch. de Psych. 1930, 22, 313-327.

35. ———and *Holzinger, K.* The Average Value for the Probable Error of Tetrad Differences. Brit. J. of Psych. 1930, 20, 368-370.

36. *Symonds, P. M.* Measurements in Secondary Education. 1929, N.Y. The McMillan Co.

37. *Thomson, G. H.* The Tetrad Difference Criterion. Brit. J. of Psych. 1927, 17, 235-255.

38. ———On the Formation of Structure Diagrams Between Four Correlated Variables. J. Ed. Psych. 1927, Vol. XVIII. No. 3.

39. *Thurstone, L. L.* Standard Tests for Office Clerks. J. App. Psych. Sept. 1929, 3, 248-251.

40. ———Multiple Factor Analysis. Conference on Individual Differences, New Haven, Conn. May 16, 1931. National Research Council, Washington, D.C.

41. *Walker, H.* Studies in the History of Statistical Method. 1929, Baltimore, The Williams and Wilkins Co.

42. *Wilson, E. B.* Comment on Professor Spearman's Note. J. Ed. Psych. 1929, 20, No. 3. 217-223.

43. *Yule, G. U.* An Introduction to the Theory of Statistics. London, Charles Griffin and Co.

APPENDIX I.

THE VALIDITY OF THE SPEARMAN METHODS.

No attempt is being made here to make any experimental or mathematical attack on the question of the validity of the Spearman methods. This is something that must be left to the statistician. The reader is entitled to ask, however, just how much reliance he can put in results obtained from the use of the Spearman methods, and the writer is attempting to give a partial answer by referring to the writings of persons interested in Spearman's theory and mathematics, and qualified to consider them.

For some years the chief opponent to Spearman's theory was Godfrey H. Thomson. His chief attack was against the conception of "g" as a single indivisible factor to be thought of in terms of mental energy. His approach to a disproof of this theory was through a dice throwing experiment where group factors (13 of them) were brought in by counting in a pre-arranged manner the number of pips into several scores. The result was that no tetrad differences over three times the probable error were found, and the conclusion was drawn by Thomson that correlations are due to complexity of group factors and not to "g." (37).

Spearman has answered Thomson's objections, and in this has received the support of Garnett. Their reply was to the effect that Thomson's findings could be equally well interpreted as due to a common factor, or "g," plus a number of specific factors (28, Appendix vii.). From a practical point of view Thomson's objection does not seem to be very important. If there are a large number of these group factors, quite a few of them can be expected to be common to any group of abilities under consideration, and, for this group, they are equivalent to "g," though, for another group, the common factors, or "g," would not be exactly the same. There are many who are not prepared to believe that the factor common to one group of tests where the tetrad criterion is satisfied is the same as the factor common to another group where the criterion is also satisfied. Wilson, for example, has the feeling that "g" is relative to the "set up." He says (42, 220). "Thus, as far as I can see, the solution for "g" depends on what the psychologist determines it shall be," and again (42, 222) "Thus we come out at the end with a result somewhat equivalent, as I understand it, to that with which Thorndike begins. He just plainly defines the measure of "g" at the outset as given by scores on his C.A.V.D."

But, it should be noted, C.A.V.D. scores would be partly made up of measurement of specific abilities.

With one exception Wilson has supported Spearman's mathematics, and Spearman is able to say, "To begin with, however, I should like to quote with lively satisfaction his (Wilson's) explicit statement that he has examined and found correct (except for misprints) not only the mathematics in my book, 'The Abilities of Man,' but also the various mathematical papers to which this book makes reference. A similar scrutiny of the mathematics in both book and references has now been carried out by the eminent British statistician, Professor A. Bowley and, as he has been good enough to tell me, with a similarly favourable issue." (29, 212). The one exception referred to above is the determination of the amount of "g" possessed by any given individual. Wilson agrees that Spearman's formula for finding this can be rigorously deduced from the tetrad equation (see p. 10), but holds that the number of unknown quantities is always one greater than the number of equations, with the result that g_x (amount of "g" possessed by a person, x.) cannot be uniquely determined. To overcome this difficulty Spearman has proposed (29, 213) the use of the further equation $r_{ag}=1$, making $g_x=m a_x$, and has pointed out that in his laboratory a value of $r_{ag}=.99$ has more than once been obtained. But, as Wilson has observed, when $r_{ag}=1$, the raw scores on the test are all that are required.

Apparently we should be very cautious about putting undue trust in any correlation with "g," even if it approaches unity, for, it will be remembered in the present investigation the correlation of the A.C.E. Psychological test with "g" found on one occasion to be .98 was revealed elsewhere as .66. Three possible explanations for this difference suggest themselves: (1) it may be due to chance fluctuation, (2) it may be due to undetected overlap, (3) the central factors in the two groups may not be the same.

Whatever the explanation, the fact remains that Spearman's methods give results that are subject to rather wide variations, even though these variations may be wholly due to the influence of chance and other forces acting on the basic inter-correlations.

Another point raised by Wilson is that, even if the tetrad criterion is not satisfied, another set of inter-correlations satisfying the criterion can be obtained from scores formed linearly from the original scores. But against this transformation of correlations Spearman argues that the process of transformation

violates certain conditions, particularly the one that none of the abilities a, b, c should overlap any other (29).

Kelley (18) has pointed out certain things that he believes to be limitations in the Spearman methods. He objects to the assumption of a normal distribution for tetrad differences (required for Spearman's probable error formula) on the ground that chance errors are known to be correlated, and, consequently, chance errors in tetrads may also be expected to be correlated (18, 13). He suggests certain weaknesses in Spearman's supplementary partial correlation procedure; his chief criticisms being in connection with the probable errors of these partial correlations, these probable errors being large and unknown.

The most important criticism, apparently, that Kelley has to level against the Spearman technique is that, in his opinion, the situations in which one is really interested are not those to which it can be usefully applied. Spearman's method shows whether or not one factor suffices to explain the correlations, but what we are more likely to be interested in is, "How many and what factors suffice." (18, 14).

Although Kelley points out what he considers to be shortcomings and special hazards, he nevertheless believes that Spearman's method has much to commend it.

Somebody having suggested to him that Professor Kelley's book, "Cross Roads in the Mind of Man," had overthrown Spearman's two factor theory, Holzinger examined Kelley's material, and came to the conclusion that, "Professor Spearman's theory has not been overthrown, but supported by Professor Kelley's data." (13, 97). Holzinger further states, "The multiple factor theory initiated by Maxwell Garnett and elaborated by Professor Kelley will undoubtedly help us to unravel some of the complex relationships existing between certain mental traits, but the writer at least sees no contradiction between the evidence presented in Professor Kelley's book and Professor Spearman's theory." (13, 97).

Professor Holzinger and Miss Frances Swineford have investigated Kelley's method of multiple factor analysis, and, in the conclusions to the report of their investigation, state, "It has been shown that Professor Kelley's seventh grade correlations may be fitted by a number of patterns varying as to number of factors, positions of coefficients, and numerical value of these constants———" "The point of this paper has been to ilustrate lack of uniqueness in pattern fitting by procedures thus far employed. Some other procedures may possibly lead to better results, but the analysis represented by Professor Kelley's work cited above is clearly inadequate." (16, 258).

Other procedures for multiple factor analysis have recently been given by Thurstone (40). He explicitly states, however, that, "The present multiple factor methods in no way contradict the Spearman two-factor methods which are very ingenious and powerful in the situations to which they apply." (40, 1). Whether Thurstone's methods will permit the unique direct determination of the factors causing a set of inter-correlations, remains to be seen, but the very attempts of Kelley, Thurstone and others suggest that Spearman's methods are not entirely satisfactory for handling group factors. Thurstone states, "The two-factor methods are not applicable to situations that involve group factors, except in indirect ways." (40, 1).

The latest work suggesting caution in the complete acceptance of Spearman's theories comes from Adams (1, 2 and 3), who gives an alternative explanation of the tendency for tetrad differences of correlation coefficients to approximate to zero. He holds that this is dependent on the identity of the standards used in measuring performance. Adams, like Spearman, bases his conclusions on experimental evidence and mathematical deduction. His mathematics parallel rather closely those of Spearman, but there are important differences that lead him to deny the validity of the division into "g" and "s" by the Spearman methods. In place of "g" Adams uses "x" which represents the standard and is always equal to 1.00, and in place of "s" he uses "e" which is always equal to zero, and is derived from uncorrelated errors. Consequently, he concludes "that any attempt to isolate two factors, such as "g" and "s" from tables of correlations which satisfy the tetrad difference criterion is futile. "g" (or "x") will be equal to 1, and "s" (or "e") will be equal to 0——— ———a, consequently, is the only useful unknown and its value can be obtained more quickly and easily from the percentage of items correct, or from the correlation of the performance with the standard than by a dissection of tables of correlations." (2, 362).

Adams treats overlapping and group factors as correlated errors, and holds that they appear only when two or more standards are being used.

So far, no reply to these alternative explanations has appeared, and, while the present writer fully expects that Spearman's theory will live through this as through the many other storms, Adams's work does seem to have sufficient merit to demand attention. Even if correlation coefficients do represent relations between the standards by which abilities are measured, we are forced to ask, "What are standards?" They are simply means of measuring ability, and, when an ability is scored in

terms of a standard, the score represents approach to perfection in that ability. We do not seem to be able to get away from the ability. Even if we have a multi-standard performance measured against a single standard scale, that is something often done in life situations, and, if the Spearman methods enable us to detect and handle to some extent such situations, they are of distinct value. And, what is a very strong argument in their favour, applications of the Spearman methods do reveal them as "working." This was seen in the present study by the agreement of overlap and the absence of it with what could be expected from observation. In his last article (3) Adams draws a conclusion which he says, "would imply that the Spearman methods are sound as methods." Perhaps it should also be pointed out that Spearman has long been aware of the effect of the standard, or at least the scale, on correlation coefficients, and, consequently, on the values found by his methods. (30).

Reference has already been made (see pages 30 and 31) to the boundary conditions claimed by Holzinger as requiring to be fulfilled before we can be sure a common factor does exist. These conditions do not prove the lack of validity in Spearman's methods, but they do give warning that invalid conclusions may be drawn, if they are not kept in mind.

Spearman himself has complete confidence in his own theory, though he is more alive to its limitations than many imagine. He has said, "such, then, is the general theory of two factors with its six foundation pillars: correlated coefficients; calculated deviations of tetrad differences from zero; observation of these deviations; proof of the two factors; their relative weights in abilities; and their actual measurement in individuals. None of these six is in the least assumptive; every one of them is a matter of rigorous demonstration. And not one of them appears at the present day to be seriously challenged by any psychologist of competence. Those who still seem to oppose them do so only by mixing them up with the "sub-theories" which seek to explain them—and which, no doubt, to introduce controversial matter." (31, 344).

Holzinger states concerning the Spearman methods: "These principles, in the opinion of the writer, constitute the most important contribution ever made in the application of statistical procedure to the data of mental life." (14, 14).

APPENDIX II.

THE RELIABILITY OF THE SPEARMAN METHODS.

The Question of Probable Errors.

The reader has undoubtedly felt the need of the addition of probable errors to the values found by the Spearman procedures. Unfortunately however, there is no known means of calculating the probable error for any one of the values arrived at in this study.

The nearest probable error is one given by Kelley (18, 41). This formula gives the probable error for the product of two correlations divided by a third, that is, for the square of the correlation of an ability with a central factor where this correlation is obtained from three tests only. However even apart from the extreme laboriousness of using this formular, it is not applicable to the methods used in this study. It was rendered invalid by the method of averaging the correlations and products of pairs of correlations. The justification for averaging these quantities is that a much more reliable value results, and this is of more use than is a value obtained from only three correlations, particularly as the probable error for the latter may be very large. Further, Kelley's formular gives the variation that could be expected from repeated applications of the same tests, while what one is more likely to be interested in is the variation likely to be found when other tests are used in the derivation of correlations with the central factor.

An empirical approach to this last problem was attempted by the writer, but without success, from the data made available by the present study. However, certain peculiarities were revealed, and these, at least on the surface, throw conflicting light on values obtained by the Spearman methods. From the correlations given in Table V all possible groups of three non-overlapping tests containing test 1 were selected, and the correlations with "g" calculated. The same thing was done for similar groups containing test 7, for groups containing test 8, and for groups containing test 3 of Table XVII. The results for test 8 of Table V are given in the table below, and these results are similar to those obtained for the other three sets.

Variations in Correlations with "g" obtained from different groups of Three Non-overlapping Tests.

Groups of Tests.	Correlations with "g."	Groups of Tests.	Correlations. with "g."
813	.79	826	.77
814	.64	827	.75
815	$(-.278)^{\frac{1}{2}}$	829	.79
816	.79	837	.47
817	.70	847	.48
819	.79	857	.47
823	1.02	867	.58
824	$(-.232)^{\frac{1}{2}}$	879	.57
825	$(-.232)^{\frac{1}{2}}$		

From the above table certain observances can be made. With four exceptions all the correlations with "g" lie in the not very wide range of .47 to .79. The four exceptions contain one correlation that is positive and greater than unity and three that are the square roots of negative quantities. The evidence regarding reliability is therefore conflicting. Disregarding the "four exceptions" quite satisfactory reliability of the Spearman method is indicated, but the "exceptions" being very wide apart and being impossible correlations the conclusion cannot be drawn that the correlation of a test with "g" is not subject to wide fluctuations. The state of affairs for the other three sets of groups of three tests is very much the same.

However, in all the "exceptions" the correlation forming the denominator is found to be very close to zero, and consequently small variations greatly affect the value obtained. Consequently, one thing does seem certain, and that is that we should not use Spearman's method where we have correlations that are very small. And, also, the evidence is quite strong that, if we use only correlations that are not less than about .20 and that are not derived from overlapping abilities, the values arrived at will not vary to any greatly disturbing extent. However, our previous conclusion (p. 58) that the use of a small number of tests is of very little value is not challenged, but instead is supported, by the data just considered.

VITA.

Christopher Jorgensen was born on August 14th, 1903, at Kerang, Victoria, Australia. His secondary school education was obtained at the Kerang and Bendigo high schools. In 1923 he entered the Melbourne Teachers' College where he obtained the Trained Primary Teachers' Certificate. Subsequent to this he studied at the University of Melbourne where the B.A. degree and the Diploma of Education were obtained. His professional experience was obtained as a junior teacher in 1922, and, after training, as a teacher in various secondary schools in the state of Victoria, Australia till August 1930, when, as holder of the R.S.S.I.L.A. Travelling Scholarship, he left to study at Columbia University.